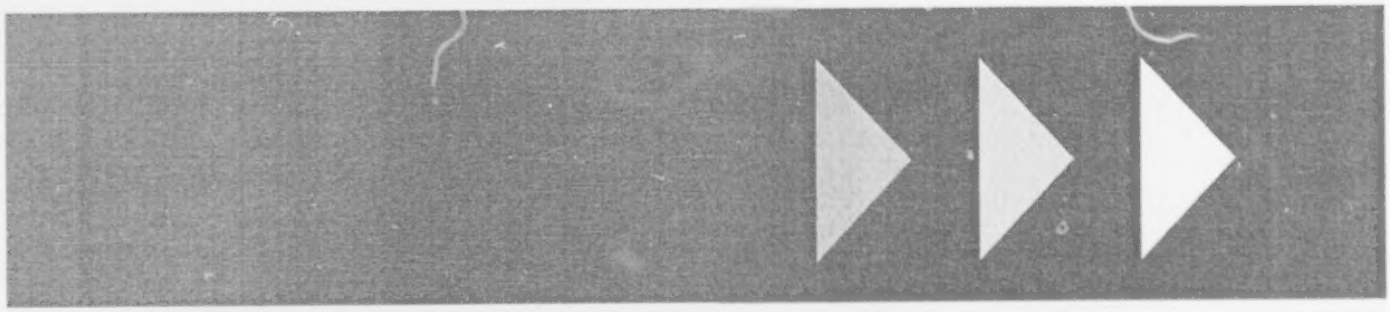
Secrétariat du Conseil du Trésor du Canada
Treasury Board of Canada Secretariat

Examen de la qualité des évaluations dans les ministères et les organismes



Canada

105-03645/

Remerciements

Un groupe de travail a été mis sur pied afin de fournir des commentaires et des suggestions pour l'examen. Nous désirons exprimer nos sincères remerciements aux personnes suivantes :

- ACIA - Theresa Iuliano
- Santé Canada - Walter Zubrycky
- Diversification de l'économie de l'Ouest Canada - Kathy Locke
- Service correctionnel du Canada - Christa Gillis
- Transports Canada - Unnati Vasavada
- DRHC - Serge Bertrand
- AAC - Eric Seraphim
- MAECI - Stephen Kester

Elles ont fourni de la rétroaction sur le mandat de l'étude, des suggestions concernant le gabarit d'examen et des commentaires sur l'ébauche du rapport.

Nous sommes très reconnaissants à Glenn Crone et Zeljka Spasojevic du Centre d'excellence en évaluation, Secrétariat du Conseil du Trésor du Canada, de leur soutien continu.

Les membres du Groupe ont travaillé en collaboration avec Shelley Borys, Michael Callahan, Janice Remai de EKOS Research Associates, Inc.

Table des matières

Résumé 1
Introduction 1
Objet 1
Méthodologie 1
Constatations 2
Conclusions et Recommandations 4

1. Introduction 5
   1.1 Politique d'évaluation du SCT 5
   1.2 Centre d'excellence en évaluation 6
   1.3 Organisation du rapport 7

2. Méthodologie 7
   2.1 Conception du Guide d'examen des évaluations 7
   2.2 Échantillon 9
   2.3 Examen des Rapports d'évaluation 11
   2.4 Analyse 11

3. Constatations 13
   3.1 Qualité des évaluations de la fonction publique fédérale : Aperçu et faits saillants 13
   3.2 Constatations détaillées 17
   3.3 Forces et faiblesses des évaluations de la fonction publique fédérale 35
   3.4 Variations de la qualité selon les caractéristiques organisationnelles et la date du rapport 39

4. Conclusions et Recommandations 44
   4.1 Conclusions 44
   4.2 Recommandations 45

Annexe A Gabarit d'examen 47

Annexe B Répartition des rapports examinés par ministère et organisme 65

Résumé

Introduction

L'évaluation soutient le but du gouvernement du Canada de devenir une organisation apprenante. Elle y réussit en aidant les cadres supérieurs, les gestionnaires de programme et les décideurs à découvrir si leurs initiatives fonctionnent ou non et si elles atteignent les objectifs; elle les aide aussi à établir s'il y a lieu de poursuivre ces initiatives et à déterminer comment mieux les concevoir et les exécuter afin d'atteindre les objectifs d'une façon rentable. La qualité des rapports d'évaluation est primordiale pour que la fonction d'évaluation réponde à ces besoins d'information.

Objet

En 2001, le Secrétariat du Conseil du Trésor (SCT) a créé le Centre d'excellence en évaluation (CEE) et a établi une nouvelle Politique d'évaluation destinée à renforcer la fonction d'évaluation et la qualité des rapports. L'un des objectifs clés du présent rapport est de déterminer si la qualité des rapports est acceptable et s'il y a eu une amélioration de la qualité. L'un des aspects importants de ce travail est de promouvoir des rapports d'évaluation de qualité. Le présent examen représente un élément de la stratégie d'ensemble du CEE pour surveiller et renforcer la qualité des rapports. Les autres activités comprennent : l'étude des meilleures pratiques; une enquête annuelle sur la santé des unités d'évaluation des ministères et des petits organismes; des rencontres individuelles, l'examen continu des évaluations, les Cadres de gestion et de responsabilisation axés sur les résultats (CGRR), les plans d'évaluation ministériels et un rapport annuel documentant les constatations des évaluations et indiquant comment elles contribuent au renforcement de la responsabilisation et à l'examen des dépenses du gouvernement.

Méthodologie

Diverses ressources ont servi à l'élaboration des critères que nous avons utilisés pour notre examen, dont le Guide d'examen des rapports d'évaluation du Centre d'excellence en évaluation, SCT, janvier 2004 et des extraits du Rapport de 1993 du Bureau du vérificateur général du Canada (BVG) sur l'évaluation de programmes (critiques des rapports d'évaluation). Nous avons

également consulté un groupe de référence d'unités d'évaluation de ministères et d'organismes. Le gabarit que nous avons utilisé pour l'examen se trouve à l'Annexe A. [1]

Constatations

Les constatations du présent l'examen indiquent que la plupart des rapports d'évaluation de la fonction publique fédérale sont de qualité acceptable, quoique nous ayons déterminé que presque le quart des évaluations (23 %) étaient inadéquates dans l'ensemble. Nous n'avons pas observé de variations claires et constantes de la qualité pour les organisations fédérales de tailles différentes et pour les ministères par rapport aux organismes. Une comparaison des rapports élaborés avant avril 2002 et de ceux qui ont été élaborés par la suite démontre toutefois une amélioration de la qualité concernant un certain nombre de critères dans les évaluations les plus récentes. Par exemple : les questions de rentabilité ont été traitées; il y a une rigueur méthodologique; les solutions de remplacement sont indiquées; les constatations présentées sont fondées sur des données; ils contiennent des recommandations formelles. Cette amélioration de la qualité avec le temps laisse croire que les efforts du SCT pour améliorer la qualité des évaluations ont peut-être un effet positif, en ayant accordé une année, jusqu'en avril 2002, aux ministères et aux organismes pour comprendre entièrement la Politique et pour donner le temps au nouveau Centre d'excellence en évaluation de commencer à fonctionner. Mais il y a un besoin pressant d'améliorer encore les choses, comme l'indiquent les constatations qui suivent.

Les principales ***forces*** des évaluations examinées dans le présent examen incluent ce qui suit :

- une description complète du programme ou de l'initiative examiné, y compris ses ressources, ses bénéficiaires et ses intéressés;
- une déclaration claire des objectifs d'évaluation;
- le recours à des sources de données multiples dans la méthodologie;
- une bonne présentation des constatations, en particulier, concernant les éléments de la pertinence, de l'exécution et de la mise en œuvre;

1 Nous avons d'abord voulu utiliser un échantillon stratifié de rapports d'évaluation en fonction de variables clés d'intérêt. Il s'est avéré que l'ensemble de rapports à prendre en considération dans notre examen était constitué seulement des rapports d'évaluation qui ont été présentés au SCT. Même si les ministères sont tenus de présenter tous les rapports d'évaluation produits, ils ne semblent pas le faire de manière fiable. Selon l'étude d'évaluation des capacités exécutée par le CEE il y a deux ans, il semble qu'environ 250 évaluations sont effectuées chaque année, ce qui aurait dû donner 500 rapports à examiner. Mais le SCT a reçu seulement 214 rapports achevés au cours des deux dernières années (les années visées par le présent examen). De plus, un grand nombre de dossiers d'évaluation sont en format électronique (liens sur le web et examens par exemple) et ne respectent pas la définition de copie complète sur papier d'une évaluation disponible aux fins d'examen. Étant donné que l'échantillon ne comprend pas l'ensemble des rapports, il est difficile d'évaluer dans quelle mesure le lot des rapports examinés est biaisé ou non. La distribution des rapports examinés par ministère et par organisme se trouve à l'Annexe B.

- des recommandations formelles ou des propositions d'améliorations qui découlent logiquement des constatations et des conclusions;
- des rapports bien rédigés et bien organisés.

Voici d'autre part quelques faiblesses des évaluations et des rapports :

- seulement six rapports d'évaluation sur dix expliquaient le choix du moment et la signification de l'évaluation;
- la plupart des rapports (les deux tiers) ne faisaient qu'énumérer les éléments d'évaluation et très peu en discutaient (environ le quart);
- les questions de rentabilité étaient couvertes de façon superficielle;
- un grand nombre de rapports ne contenaient pas de description complète des principaux aspects de la méthodologie; si juste un peu plus de la moitié des rapports décrivaient la méthodologie, quatre sur dix ont seulement énuméré quelques aspects de la méthodologie et seulement le quart ont fait référence à un document technique;
- on note peu d'intégration de données issues d'un système de mesure du rendement;
- seulement une minorité des concepts d'évaluation incluaient des caractéristiques d'optimisation de la rigueur de l'analyse comme un groupe de comparaison (13 %), des mesures de référence (14 %) ou une comparaison avec des normes, de la documentation ou un autre point de référence (22 %). Seulement 26 % ont utilisé des entrevues avec des répondants indépendants clés qui n'avaient pas un intérêt dans le programme;
- seulement environ quatre rapports d'évaluation sur dix comprenaient une déclaration des limites ou des contraintes de l'évaluation;
- seulement environ le tiers des évaluations présentaient des constatations permettant de déterminer si le programme fait double emploi ou s'il est en contraste avec d'autres programmes ou d'autres initiatives;
- seulement le quart des évaluations examinaient les résultats fortuits (25 %) ou traitaient les impacts différentiels (26 %);
- seulement 26 % des évaluations présentaient des constatations concernant des démarches de remplacement susceptibles d'être plus rentables, quoique la couverture de cet élément ait augmenté dans les rapports les plus récents (31 % après avril 2002 contre 16 % avant avril 2002);

- nous avons déterminé que presque le quart des évaluations (24 %) étaient inadéquates dans leur prestation de conclusions objectives, fondées sur des données concernant la pertinence, la réussite ou la rentabilité, ou tout cela;
- parmi les rapports qui contenaient des recommandations, seulement 26 % indiquaient des scénarios de remplacement; et
- moins de la moitié des rapports d'évaluation incluaient une réponse de la direction (48 %) ou un plan d'action (33 %);
- 25 % des rapports qui contenaient des recommandations incluaient une recommandation relative au financement global, et dans tous ces cas, la recommandation était d'accroître le financement;
- aucun rapport ne présentait de données indiquant qu'un programme n'était pas pertinent ou qu'il n'était pas nécessaire.

Conclusions et Recommandations

Dans l'ensemble, la plupart des évaluations que nous avons examinées sont de qualité raisonnable. Nous avons déterminé que la majorité étaient adéquates dans l'ensemble (45 %) ou plus qu'adéquates (32 %). Mais nous avons aussi déterminé qu'une proportion considérable des évaluations (23 %) étaient inadéquates, une constatation qu'il faut traiter. C'est pourquoi le présent rapport recommande que le Centre d'excellence en évaluation du SCT prenne les mesures suivantes :

❶ Encourager les divisions des évaluations des ministères et des organismes fédéraux à renforcer leurs rapports d'évaluation en traitant les principales faiblesses déterminées dans le présent examen.

❷ Peaufiner les lignes directrices et les critères du Conseil du Trésor concernant les caractéristiques attendues des (1) méthodologies d'évaluation et (2) des rapports d'évaluation, et les diffuser.

❸ Poursuivre la mise en oeuvre d'une démarche rigoureuse de surveillance de la qualité des évaluations et l'utiliser comme base d'élaboration de fiches de rendement individuelles sur la qualité et la santé globale de la fonction d'évaluation par ministère et par petit organisme.

❹ Indiquer des mesures, y compris une structure de stimulants et des normes, afin d'assurer que les ministères et les organismes présentent des évaluations et des examens réalisés d'une façon responsable et raisonnable. Le respect de ces normes par les ministères et les organismes devrait être annoncé publiquement.

1. Introduction

L'évaluation soutient le but du gouvernement du Canada de devenir une organisation apprenante. Elle y réussit en aidant les gestionnaires de programme et les décideurs à découvrir si leurs initiatives fonctionnent ou non et si elles atteignent les objectifs, à déterminer s'il y a lieu de poursuivre ces initiatives et à déterminer comment mieux les concevoir et les exécuter afin d'atteindre les objectifs d'une façon rentable. Le Secrétariat du Conseil du Trésor (SCT) a introduit la Politique d'évaluation (la Politique) en avril 2001 afin de clarifier le rôle important de l'évaluation dans son cadre de gestion.

Le Centre d'excellence en évaluation (CEE) a été établi en 2001 pour faciliter la mise en oeuvre de la nouvelle Politique d'évaluation et pour en surveiller la réussite. Le CEE, en surveillant les pratiques d'évaluation des ministères et des organismes fédéraux, a déterminé qu'il y avait lieu d'examiner le niveau de qualité des évaluations de ces ministères et de ces organismes, afin de déterminer les forces et les faiblesses des pratiques d'évaluation et les réponses convenables. Le présent document contient le projet de Rapport final de cet examen des évaluations de la fonction publique fédérale.

1.1 Politique d'évaluation du SCT

Dans le contexte de renouveau du gouvernement fédéral, les évaluations ont pris considérablement plus d'importance, mais il n'en a pas été ainsi de la capacité de les exécuter[2]. Les ressources, humaines et autres, consacrées à l'évaluation ont diminué régulièrement depuis le début des années 1990. De plus, la Politique d'évaluation courante a accru la tâche requise pour faire les évaluations.

La Politique d'évaluation du SCT, qui a été révisée la dernière fois le 1er avril 2001, soutient « l'engagement permanent du gouvernement en vue de l'amélioration continue de la gestion et de la responsabilisation », comme l'a déclaré la ministre Robillard dans un communiqué de presse du 14 février 2003[3]. Dans la Politique d'évaluation courante, l'évaluation a un rôle clé à jouer pour soutenir la gestion des résultats dans la fonction publique. La Politique repose sur les trois principes suivants : les gestionnaires de la fonction publique ont la responsabilité de réaliser les objectifs et d'en faire rapport; l'évaluation rigoureuse et objective est un outil important de la gestion des résultats; et les ministères et les organismes sont responsables, avec le soutien du SCT, d'assurer la rigueur des évaluations. L'objectif déclaré de la Politique est de « faire en sorte que le gouvernement dispose en temps voulu de renseignements stratégiques, objectifs et solides

2 Secrétariat du Conseil du Trésor du Canada (septembre 2003). *Politique d'évaluation : Cadre de gestion et de responsabilisation axé sur les résultats (CGRR)*. page 4. En ligne : http://www.tbs-sct.gc.ca/eval/tools_outils/polrmaf-polcgrr-PR_f.asp?printable=True

3 Politique d'évaluation : Cadre de gestion et de responsabilisation axé sur les résultats (CGRR), opt. cit.

au sujet du rendement de ses politiques, programmes et initiatives, pour produire de meilleurs résultats pour les Canadiens et les Canadiennes. » Ses exigences clés sont les suivantes :

- Établir une capacité d'évaluation adéquate, incluant la haute direction.
- Accroître la portée des évaluations en traitant notamment les politiques, les programmes et les initiatives, en plus des activités semblables qui sont exécutées dans le cadre de mécanismes de partenariat (interministériels, intergouvernementaux, etc.).
- Accorder une importance accrue à la surveillance du rendement et aux résultats initiaux. Pour ce faire :
    - Élaborer des cadres de gestion et de responsabilisation axés sur les résultats (CGRR) à l'égard des politiques, des initiatives et des programmes nouveaux ou renouvelés.
    - exécuter des activités continues de surveillance et de mesure du rendement;
    - traiter les questions liées aux activités initiales de mise en oeuvre et d'administration;
    - traiter les questions liées à la pertinence, aux résultats et à la rentabilité.
- Élaborer des plans d'évaluation stratégique.
- Incorporer l'évaluation à la gestion et à la prise de décisions stratégiques.
- Mettre en œuvre des normes de pratique simplifiées et consolidées.

1.2 Centre d'excellence en évaluation

Le CEE a été établi en même temps que la Politique d'évaluation pour assurer un leadership et faciliter la mise en œuvre de la Politique. L'examen courant de la qualité des évaluations soutiendra le mandat de surveillance et de rapport du CEE concernant l'état de la capacité d'évaluation dans l'ensemble de la fonction publique fédérale. Le CEE a été conçu pour offrir les fonctions clés suivantes :

- servir de plaque tournante du leadership de la fonction d'évaluation de la fonction publique fédérale;
- lancer des initiatives relatives à des défis partagés dans la collectivité, comme la conception d'un cadre des ressources humaines visant les besoins à long terme en matière de recrutement, de formation et de perfectionnement;

- soutenir le renforcement de la capacité, l'amélioration des pratiques et le renforcement de la collectivité de l'évaluation de la fonction publique fédérale.

Le CEE exécute à ces fins des activités comme : la mise en œuvre de politiques; la surveillance; le renforcement de la capacité; les conseils stratégiques; les communications et le réseautage.

1.3 Organisation du rapport

Le présent document contient les résultats de l'« Examen de la qualité des évaluations dans les ministères et les organismes ». Notre méthodologie est présentée au chapitre suivant. Les constatations sont présentées au chapitre 3 et les conclusions au chapitre 4.

2. Méthodologie

Le présent chapitre décrit la démarche méthodologique du projet. La description est divisée en quatre sections : la conception du guide d'examen; l'échantillon; l'examen des rapports d'évaluation; et une note concernant l'analyse.

2.1 Conception du Guide d'examen des évaluations

Nous avons évalué diverses ressources dans l'élaboration des critères aux fins du présent examen. En cherchant les indicateurs possibles de qualité pour lesquels nous allions recueillir des données pour notre examen, nous nous sommes d'abord tournés vers le Cadre de gestion et de responsabilisation axé sur les résultats (CGRR) pour la Politique d'évaluation du Secrétariat du Conseil du Trésor. Une analyse du CGRR a révélé que l'examen facilitera particulièrement le traitement du groupe des questions présentées à la Section E, Questions liées au progrès et à la réussite, c'est-à-dire : « La fonction d'évaluation produit-elle des commentaires opportuns et efficaces intégrés à la prise de décisions des ministères? »; contribuant aux résultats immédiats attendus de la Politique, à savoir *des rapports crédibles en temps voulu et fondés sur des données*. Toutefois, à cause de l'ampleur du projet, il n'est pas possible d'évaluer si les rapports sont présentés en temps voulu. En outre, nous avons examiné seulement les rapports d'évaluation achevés *depuis* que la Politique a été mise en œuvre; nous n'avons donc pas de mesure de référence de la qualité pour comparer les résultats de notre examen.

De nombreux documents traitent la question des critères de qualité qui ont été consultés au cours de la conception du présent travail. En voici quelques-uns :

- « Guide d'examen des rapports d'évaluation », élaboré par le Centre d'excellence en évaluation, SCT, janvier 2004;

- « Liste de vérification pour le contrôle interne de l'étude d'évaluation : réalisations et rapports, processus et travail des entrepreneurs », préparée par la Direction générale de l'évaluation des programmes, DRHC, septembre 2003;
- « Guide d'évaluation des rapports d'évaluation de Santé Canada », préparé par la Division de l'évaluation des programmes du Ministère, Santé Canada, avril 2003;
- un cadre d'évaluation de la qualité des évaluations, élaboré par un consultant externe à l'intention du Bureau du vérificateur général (non mis en œuvre);
- extraits du Rapport de 1993 du BVG sur l'évaluation de programmes (« critiques des rapports d'évaluation »), élaborés par le CEE.

Voici les questions centrales de notre étude : Est-ce que la qualité des rapports est acceptable et est-ce qu'il y a eu une amélioration de la qualité? Notons qu'un simple examen des rapports d'évaluation ne nous a pas permis de déterminer s'il y a eu une amélioration de la qualité des rapports. On peut obtenir ce genre d'information seulement en faisant des comparaisons avec les évaluations faites avant l'adoption de la Politique et en interrogeant les responsables. Toutefois, selon un examen de la Politique d'évaluation (y compris l'Annexe B de la Politique), de son CGRR et du matériel mentionné ci-dessus, les indicateurs possibles qui ont été identifiés pour mesurer la qualité des rapports d'évaluation comportent les caractéristiques suivantes :

- ils sont rédigés avec clarté, concision et simplicité;
- ils décrivent clairement le programme, la politique ou l'initiative examiné, y compris ses objectifs, ses extrants, ses résultats attendus, sa portée et ses ressources;
- ils ont une évaluation des résultats réalisés par la politique, le programme ou l'initiative;
- ils ont une description de l'évaluation, notamment le choix du moment; la méthodologie; les objectifs et les éléments d'évaluation; et comment l'évaluation se situe, et son importance, par rapport aux opérations générales du ministère ou de l'organisme;
- ils indiquent les limites de l'évaluation, concernant le contexte, la portée, les méthodes et les conclusions;
- ils ont une méthodologie convenable (p. ex. des sources de données multiples);
- ils ont des conclusions qui traitent clairement les principaux éléments d'évaluation concernant la pertinence, les réussites et les incidences et la rentabilité (selon le type d'évaluation – formative ou sommative);

- ils contiennent seulement l'information nécessaire pour comprendre les constatations, les conclusions et les recommandations;
- les constatations présentées sont crédibles et fondées sur des données, par exemple :
    - des données réunies dans des enquêtes auprès d'un groupe représentatif de participants et comparés à un groupe comparable de non-répondants;
    - des données dérivées de comparaisons avec des mesures de référence du système de mesure du rendement; et
    - des données qualitatives obtenues de répondants clés qui n'ont pas un intérêt dans le programme ou qui connaissent vraiment la question;
- ils ont des conclusions et des recommandations qui émanent logiquement des constatations de l'évaluation;
- ils ont des recommandations claires et atteignables indiquant les mesures à prendre et le cadre temporel;
- ils analysent et expliquent l'exposition au risque des problèmes déterminés et par rapport aux recommandations présentées.

Nous nous sommes fondés sur notre analyse de tout le matériel de référence mentionné ci-dessus pour élaborer un projet de gabarit d'examen. Comme suite à l'élaboration d'un projet d'instrument contenant les critères proposés et l'examen de celui-ci avec les autorités responsables du projet, nous avons rencontré le groupe de travail du CEE (représentant huit ministères fédéraux) pour discuter des critères et de la portée de l'examen. Les révisions apportées au gabarit d'examen tiennent compte des commentaires que nous avons reçus à ce moment-là. Le gabarit final utilisé pour l'examen se trouve à l'Annexe A.

2.2 Échantillon

Nous avions proposé de choisir l'échantillon de rapports d'évaluation dans une base de données compilée par le CEE sur les rapports d'évaluation effectués depuis l'adoption de la Politique d'évaluation, c'est-à-dire l'exercice financier 2001-2002. La « population » des rapports serait stratifiée selon certaines variables clés d'intérêt. Les titres des rapports seraient choisis en proportion des caractéristiques de la population, ou en nombres suffisants pour assurer la représentation de tous les sous-groupes clés.

Dans la mesure où la stratification était possible ou désirée, ou les deux, nous avions un certain nombre de stratifications possibles de l'échantillon et de variables de sélection, par exemple : le type d'évaluation, formative ou sommative; la taille et le type de ministère ou d'organisme; l'année de l'évaluation (on peut s'attendre en effet à ce que la qualité des évaluations augmente avec le temps, à mesure que la Politique est mise en oeuvre et que les évaluateurs et les responsables du CEE se familiarisent avec elle).

Il s'est avéré que la population des rapports à prendre en considération pour notre examen était constituée seulement des rapports d'évaluation[4] qui ont été présentés au SCT. Même si les ministères sont tenus de présenter tous les rapports d'évaluation produits, ils ne semblent pas le faire de manière fiable. Selon l'étude d'évaluation des capacités exécutée par le CEE il y a deux ans, il semble qu'environ 250 évaluations sont effectuées chaque année, ce qui aurait dû donner 500 rapports à examiner. Mais le SCT a reçu seulement 214 rapports produits au cours des deux dernières années (les années visées par le présent examen). De plus, d'autres dossiers d'évaluation sont en format électronique (liens sur le Web et examens par exemple), mais ne respectaient pas la définition de « copie complète sur papier d'une évaluation disponible aux fins d'examen ».

Étant donné le temps limité pour effectuer l'examen, il était impossible d'obtenir la série complète des rapports d'évaluation des ministères individuels. Il est de plus difficile de déterminer quelle aurait été l'incidence sur l'objectivité de l'échantillon si nous avions réclamé des ministères et des organismes qu'ils nous présentent des rapports aux fins du présent examen.

Il est donc important de noter que nous avons examiné cet échantillon limité de rapports d'évaluation qui ont été présentés au SCT et qui étaient complets. Étant donné que l'échantillon ne comprend pas l'ensemble de la population, il est difficile d'évaluer dans quelle mesure le lot des rapports examinés est biaisé ou non.

Dans le processus de recherche de rapports pour l'examen, nous avons consulté la série complète des rapports présentés après le 1er avril 2001 et offerts par l'entremise du SCT. Même si la base de données indiquait que plus de 200 rapports étaient disponibles pour cet exercice, il a été établi qu'un grand nombre des dossiers ne convenaient pas pour l'examen. Certains dossiers contenaient par exemple seulement un résumé d'un rapport, ou étaient des rapports sur des vérifications, des études spéciales (par exemple pour fournir une analyse approfondie d'une question destinée à faire partie d'une évaluation, mais qui n'était pas une évaluation en elle-même) ou d'autres types d'examen qui ne constituaient pas une évaluation.

4 Les rapports de la population et de notre échantillon (n=115) incluaient à la fois les rapports d'évaluation qui étaient obligatoires et ceux qui ne l'étaient pas. Les évaluations obligatoires (c'est-à-dire celles qui ont été faites pour appuyer une demande de renouvellement de financement de programme au Conseil du Trésor) sont axées sur des questions particulières (p.ex. celles qui sont précisées dans le CGRR) et les directives du CT sont donc claires quant à ce que ces rapports devraient traiter. Par contre, le champ des évaluations non obligatoires peut être plus étroit ou plus large, selon leur raison-d'être.

Le plan de travail consistait à examiner un total de 110 rapports. Nous disposions finalement de 122 rapports à examiner et nous en avons examiné 115. Ceux qui n'ont pas été examinés (n=7) étaient des rapports de ministères qui étaient déjà très représentés dans l'échantillon. Nous avons essayé de limiter le nombre total de rapports examinés pour un ministère en particulier, afin d'assurer la représentation dans la population des rapports disponibles. Il s'est avéré que plusieurs ministères avaient 10 ou 12 rapports qui ont été examinés (et ces ministères étaient également ceux qui offraient des rapports qui n'ont pas été examinés).

Le CEE avait élaboré six des rapports de l'échantillon. Comme il n'était pas convenable que nous examinions nos propres rapports, des analystes du SCT ont été formés pour appliquer le gabarit d'examen et ont ensuite exécuté les examens de cinq de ces six rapports (le sixième rapport provenait d'un ministère déjà bien représenté et n'était par conséquent pas nécessaire).

La distribution des rapports examinés par ministère et par organisme se trouve à l'Annexe B.

2.3 Examen des Rapports d'évaluation

Un long processus prétest auquel ont pris part tous les examinateurs a été entrepris, non seulement pour mettre à l'épreuve le gabarit d'examen, mais également pour assurer la fiabilité inter-évaluateur. Chacun des membres de l'équipe de base a examiné un total de trois rapports. Après l'examen de chaque rapport et l'application du gabarit, l'équipe s'est réunie pour discuter en détails des notes attribuées par chaque examinateur. Quand il y avait des écarts, la discussion subséquente a permis de clarifier la signification de certains aspects de l'examen ou de certaines notes. Le gabarit a également été révisé pour tenir compte de ces éclaircissements additionnels quand c'était possible. On utilisait ensuite le gabarit révisé pour l'examen de prétest suivant. Il a été déterminé à la fin du troisième examen de prétest que la fiabilité inter-évaluateur (évaluée qualitativement) était suffisamment élevée pour entreprendre les examens indépendants.

Suite aux prétests et à la mise au point du gabarit d'examen, nous avons entrepris l'examen complet des évaluations. Chacun des rapports d'évaluation a été évalué par un seul examinateur. Tous les examinateurs étaient des évaluateurs chevronnés possédant des antécédents considérables en matière d'évaluation des programmes de la fonction publique fédérale. L'examen de chaque rapport a duré en moyenne deux heures et demie.

2.4 Analyse

Nous avons exécuté des analyses à une variable et des analyses croisées sur les données des examens. La plupart des critères évalués dans les examens ont été notés sur une échelle de cinq points allant de 1 (« médiocre ») à 5 (« excellente »), la note médiane 3 signifiant « adéquate ». Nous avons fait une réduction pour les analyses, les notes y représentant alors les trois catégories suivantes : 1-2 (« inadéquate »), 3 (« adéquate ») et 4-5 (« plus qu'adéquate »). Nous avons fait

ensuite des croisements selon la taille du ministère ou de l'organisme. Trois catégories ont été élaborées : petite (500 ETP ou moins, n=18) [5]; moyenne (501 à 4 600 ETP, n=51); et grande (plus de 4 600 employés, n=46). Nous avons également fait des croisements sur l'année du rapport (jusqu'à mars 2002, n=37, et avril 2002 et après, n=78) et également sur les ministères (n=91) par rapport aux organismes (n=24). Les tableaux des résultats sont présentés à l'Annexe technique dans un document distinct.

A) Limites

La qualité des évaluations peut être mesurée de différentes manières. Durant cet examen, nous nous sommes penchés sur la qualité des évaluations, selon les rapports d'évaluation. Il convient de signaler qu'une autre dimension importante de la qualité des évaluations, non examinée durant cet examen, est l'utilité, comme l'indique le degré de mise en œuvre des recommandations d'évaluation. Le CEE a précisé qu'il examinera ce critère de qualité en recourant à d'autres sources de données.

Il est important de noter qu'à titre d'examinateurs externes d'un rapport d'évaluation, nous n'avons pas toujours eu toute l'information sur les limites possibles d'une évaluation particulière (p.ex., les restrictions budgétaires, le temps alloué, les contraintes internes) ou le contexte (nous n'avons pas interrogé les gestionnaires de programme ou des évaluations). Il est donc possible que certains rapports apparaissent faibles dans notre examen, même s'ils étaient peut-être en fait très forts compte tenu du contexte ou des limites externes qu'ils comportaient.

Le Groupe de travail du CEE a aussi mentionné que la qualité des rapports d'évaluation sur lesquels a porté cet examen peut sembler faible à certains égards parce que les ministères ignoraient les critères d'évaluation avant d'effectuer les évaluations. De plus, il a mentionné que les ministères peuvent avoir évalué ou inclus des critères prévus dans cet examen, sans en traiter dans le rapport d'évaluation, parce que ceux-ci ont fait l'objet d'autres rapports ou n'ont pas été jugés pertinents pour ce rapport.

De plus, en raison des contraintes budgétaires et temporelles qui s'appliquaient au présent examen (seulement deux heures et demie pour examiner chaque rapport), il a été déterminé avec le client au cours de la phase de conception que l'examen serait essentiellement quantitatif (c'est-à-dire des articles fermés dans le gabarit d'examen présenté à l'Annexe A). Par conséquent, l'information qualitative détaillée expliquant les diverses notes de chaque rapport d'évaluation n'a pas été recueillie.

---

5 Le petit nombre de rapports de petits organismes fait en sorte que les constatations relatives à cette catégorie doivent être traitées avec précaution.

3. Constatations

3.1 Qualité des évaluations de la fonction publique fédérale : Aperçu et faits saillants

Les constatations du présent examen indiquent que la plupart des rapports d'évaluation de la fonction publique fédérale sont de qualité acceptable, quoique nous ayons déterminé que presque le quart des évaluations (23 %) étaient inadéquates dans l'ensemble. Nous n'avons pas observé de variations claires et constantes de la qualité pour les organisations fédérales de tailles différentes et pour les ministères par rapport aux organismes. Une comparaison des rapports produits avant avril 2002 et de ceux produits par la suite démontre toutefois une amélioration de la qualité concernant un certain nombre de critères dans les évaluations les plus récentes. Cela laisse croire que la Politique d'évaluation d'avril 2001 du SCT a peut-être eu un effet positif (c'est-à-dire en accordant une année, jusqu'en avril 2002, aux ministères et aux organismes pour comprendre entièrement la Politique et avoir le temps de mettre en œuvre quelques améliorations). Mais il y a un besoin d'améliorer encore les choses, comme l'indiquent les faiblesses mentionnées ci-dessous.

L'examen révèle que les rapports d'évaluation de la fonction publique fédérale ont un certain nombre de forces et de limites, dont il n'y a pas de modèle clair (c'est-à-dire qu'une section donnée des rapports, comme l'introduction et le contexte, contient à la fois des forces et des faiblesses selon le critère particulier évalué). Les principales forces des évaluations examinées dans le présent examen incluent :

- une description complète du programme ou de l'initiative examiné, y compris ses ressources, ses bénéficiaires et ses intéressés;
- une déclaration claire des objectifs d'évaluation;
- le recours à des sources de données multiples dans la méthodologie;
- une bonne présentation des constatations, en particulier, concernant la pertinence et les éléments d'exécution et de mise en œuvre;
- des recommandations formelles ou des propositions d'améliorations qui découlent logiquement des constatations et des conclusions;
- des rapports bien rédigés et bien organisés.

Voici d'autre part quelques faiblesses des évaluations et des rapports :

- absence de modèle de logique du programme ou de référence;
- un examen inadéquat des éléments d'évaluation et l'absence des références des documents sources comme les CGRR ou les Cadres d'évaluation;
- une description inadéquate des aspects de la méthodologie et des instruments de collecte des données qui ne sont pas annexés ou dont la référence n'est pas indiquée;
- l'utilisation inadéquate des données de surveillance du rendement et des points de vue des répondants clés indépendants qui n'ont pas un intérêt dans le programme;
- une évaluation inadéquate de l'impact différentiel des programmes et, par conséquent, l'utilisation insuffisante de groupes de comparaison et de mesures de référence dans la conception des évaluations;
- une couverture superficielle des questions de rentabilité.

Voici les faits saillants des constatations concernant chacune des grandes questions et des grands éléments évalués dans l'examen :

- **Résumé :** Même si la plupart des rapports (86 %) incluaient un résumé, ce dernier a besoin d'être amélioré. Nous avons déterminé que le quart des résumés que nous avons examinés étaient inadéquats[6] à titre de document cohérent et autonome et qu'environ le tiers ne présentaient pas du tout les éléments d'évaluation – quoique ce dernier défaut soit moins courant dans les rapports présentés après avril 2002 (22 %) que dans ceux présentés avant (56 %).
- **Introduction et contexte :** La plupart des rapports d'évaluation examinés présentaient bien le programme ou l'initiative évalué, y compris ses ressources, ses bénéficiaires et ses intéressés. De plus, environ six rapports sur dix traitaient des hypothèses sous-jacentes du programme (comme le financement, les partenariats), de facteurs externes comme les influences environnementales, du moment choisi pour l'évaluation et de son importance. La plupart des rapports incluaient aussi une déclaration claire des objectifs de l'évaluation. D'autre part, la plupart des rapports ne présentaient pas de modèle logique ou n'en indiquaient pas la référence et n'examinaient pas les grands rapports de cause et d'effet sur lesquels le programme était basé (moins du quart des évaluations incluaient ces éléments). La

6 La plupart des critères évalués dans le présent examen ont été notés sur une échelle de cinq points allant de 1 (« médiocre ») à 5 (« excellent »), la note médiane 3 signifiant « adéquat ». En présentant les constatations dans le présent chapitre, nous avons fait une réduction pour que les notes représentent les trois catégories suivantes : 1-2 (« inadéquat »), 3 (« adéquat ») et 4-5 (« plus qu'adéquat »).

plupart des rapports (les deux tiers) ne faisaient qu'énumérer les éléments d'évaluation et très peu (environ le quart) en discutaient. De plus, la moitié des rapports ne faisaient référence à aucun document, comme un CGRR ou un Cadre d'évaluation, comme contexte d'élaboration des éléments d'évaluation.

- **Méthodologie :** La majorité des évaluations (72 %) ont utilisé un concept de recherche convenable, à la lumière des objectifs de l'étude. Nous avons constaté que seulement 5 % n'avaient pas été conçues adéquatement (par exemple parce que très peu de répondants ont été consultés ou parce que la gamme des perspectives incluses était limitée), mais c'est un critère que nous avons été incapables d'évaluer pour presque le quart (23 %) des rapports, à cause du manque de détails. Parmi les rapports évalués, nous avons déterminé que la qualité de la conception de la méthodologie était adéquate ou mieux pour 87 % des évaluations. À peu près toutes les évaluations (97 %) incluaient des sources de données multiples. Mais il y avait aussi des faiblesses. Un grand nombre de rapports ne contenaient pas de description complète des principaux aspects de la méthodologie. Si juste un peu plus de la moitié des rapports décrivaient la méthodologie, quatre rapports sur dix en ont seulement énuméré quelques aspects. Seulement le quart des rapports ont fait référence à un document technique plus étoffé. Nous avons déterminé par conséquent que 46 % des rapports étaient inadéquats du point de vue de leur description méthodologique. De plus, la moitié des rapports ne contenaient pas d'instruments de collecte des données ou une référence indiquant où trouver les instruments. Seulement une minorité des évaluations comprenaient des données d'un système de mesure du rendement (24 %) ou des données d'entrevue avec des répondants indépendants clés qui n'avaient pas un intérêt dans le programme (26 %). Cette dernière caractéristique est toutefois plus courante dans les évaluations faites après avril 2002 que dans celles faites avant (31 % contre 16 %). Seulement une minorité des concepts d'évaluation incluaient un groupe de comparaison (13 %), des mesures de référence (14 %) ou une comparaison avec des normes, de la documentation ou un autre point de référence (22 %), soit des caractéristiques qui peuvent améliorer la rigueur de la méthodologie. Finalement, seulement environ quatre rapports d'évaluation sur dix comprenaient une déclaration des limites ou des contraintes de l'évaluation.

- **Constatations – Pertinence :** Plus de la moitié des évaluations (juste un peu moins de 60 %) présentaient des constatations relatives à la nécessité de poursuivre le programme et à sa pertinence. Nous avons déterminé que la majorité (85 %) de ces évaluations étaient adéquates ou plus qu'adéquates en fonction de ces critères. Seulement environ le tiers des évaluations présentaient des constatations permettant de déterminer si le programme faisait double emploi ou s'il était en contraste avec d'autres programmes ou d'autres initiatives; quand elles en présentaient, nous avons déterminé que ces présentations étaient inadéquates dans 18 % des cas.

- **Constatations – Réussites :** La majorité des évaluations (87 %) ont présenté des constatations démontrant si le programme ou l'initiative en question produisait ou non des résultats appuyant sa poursuite ou son renouvellement. Même si nous avons déterminé qu'environ le quart de ces rapports (26 %) étaient inadéquats en fonction de ce critère, la proportion dont la présentation de ces résultats était moins qu'adéquate a diminué (19 % après avril 2002 contre 39 % avant avril 2002). Seulement le quart des évaluations examinaient les résultats fortuits (25 %) ou traitaient les impacts différentiels (26 %). Environ les deux tiers des évaluations n'ont pas traité ces éléments.

- **Constatations – Rentabilité :** Seulement 26 % des évaluations présentaient des constatations concernant des démarches de remplacement susceptibles d'être plus rentables, quoique la couverture de cet élément ait augmenté dans les rapports les plus récents (31 % après avril 2002 contre 16 % avant avril 2002). De plus, environ le tiers des évaluations (34 %) offraient une évaluation qualitative, ou quantitative, ou les deux, de la rentabilité du programme ou de l'initiative examiné, quoique nous ayons déterminé que 28 % de ces évaluations étaient inadéquates en fonction de ce critère.

- **Constatations – Exécution et mise en oeuvre** En ce qui a trait aux éléments de l'exécution et de la mise en œuvre, la plupart des évaluations présentaient des constatations relatives à la pertinence du modèle d'exécution du programme, ou des pratiques de gestion, ou les deux (81 %), et à la nécessité d'améliorer la structure du programme ou les ententes d'exécution (76 %). Les évaluations étaient très bien notées en fonction du premier critère (89 % étaient adéquates ou plus qu'adéquates).

- **Constatations – Pertinence de l'analyse :** Il était difficile d'évaluer la pertinence de l'analyse (c'est-à-dire la mesure dans laquelle l'analyse était appuyée par les données selon les tests de signification, les taux de réponse, etc.) pour 50 % des évaluations en raison du manque de détails présentés dans les rapports. Nous avons déterminé que presque le tiers (32 %) des rapports examinés en fonction de ces critères étaient inadéquats. Cette dernière proportion a toutefois diminué au cours des dernières années (26 % après avril 2002 par rapport à 41 % avant avril 2002).

- **Conclusions :** Nous avons déterminé que les trois quarts des évaluations étaient adéquates ou mieux et que le quart (24 %) étaient inadéquates en ce qui a trait à leur prestation de conclusions objectives, fondées sur des données concernant la pertinence, la réussite ou la rentabilité, ou tout cela. Parmi les évaluations qui traitaient la mise en œuvre, l'exécution ou les pratiques de gestion, ou tout cela, nous avons déterminé qu'une proportion plus élevée (85 %) étaient adéquates ou mieux dans la prestation de conclusions objectives, fondées sur des données concernant ces éléments. De plus, la qualité des évaluations s'améliore concernant ce critère : nous avons déterminé que 40 % des évaluations faites après avril 2002 étaient plus qu'adéquates à cet égard par rapport à seulement 20 % des rapports produits

avant cette date. De plus, dans leurs conclusions, la moitié des évaluations (49 %) présentaient d'autres leçons apprises au sujet du programme. Nous avons déterminé que 95 % de ces rapports étaient adéquats ou plus qu'adéquats à cet égard.

- **Recommandations :** La grande majorité des évaluations incluaient des recommandations formelles (77 %) ou des suggestions de mesures à prendre (13 %). Dans presque tous les cas, les recommandations traitaient des constatations d'évaluation importantes et découlaient logiquement des constatations et des conclusions (94 % dans chaque cas). D'autre part, parmi les rapports qui contenaient des recommandations, seulement 26 % indiquaient des scénarios de remplacement et seulement 35 % tenaient compte des contraintes pratiques (par exemple les règlements, les budgets). Nous avons déterminé que plus du tiers de ces rapports (35 %) étaient inadéquats en fonction de ce critère.

- **Réponse de la direction et plan d'action :** Moins de la moitié des rapports d'évaluation incluaient une réponse de la direction (48 %) ou un plan d'action (33 %).

- **Aspects généraux et autres du rapport :** Nous avons déterminé que la plupart des rapports d'évaluation étaient adéquats ou plus qu'adéquats quant à la clarté de la rédaction (86 %) et à l'organisation (81 %). En ce qui a trait aux faiblesses, nous avons déterminé qu'une proportion considérable des rapports étaient inadéquats concernant la présentation acceptable des données, y compris les chiffres et les sources (33 %), la présentation convenable de l'information technique (30 %) et l'utilisation efficace des tableaux et des graphiques (25 %).

- **Évaluation globale :** La majorité des rapports d'évaluation ont reçu une note globale subjective indiquant qu'ils étaient adéquats (45 %) ou plus qu'adéquats (32 %), même si nous avons déterminé que presque le quart des évaluations (23 %) étaient inadéquates.

3.2 Constatations détaillées

A) Résumé

La majorité des rapports examinés (86 %) comprenaient un résumé. Les ministères étaient plus susceptibles d'inclure un résumé dans leurs rapports d'évaluation que les organismes (90 % contre 71 %). De plus, les organisations de grande et de moyenne tailles (83 % et 92 % respectivement) étaient plus susceptibles d'inclure un résumé que les petites organisations (78 %).

En ce qui a trait à la clarté et à la concision de la rédaction et à la cohérence à titre de document autonome, nous avons déterminé que la plupart des résumés étaient adéquats ou plus qu'adéquats (43 % et 31 % respectivement) et que le quart étaient inadéquats.

Autres caractéristiques clés des résumés :

- Les éléments d'évaluation clés étaient présentés de façon complète (38 %) ou partielle (30 %) dans la plupart des résumés, mais pas du tout dans 32 % des résumés des rapports. Les résumés sans présentation des éléments d'évaluation étaient plus courants dans les rapports présentés avant avril 2002 que dans ceux qui ont été présentés après (56 % contre 22 %) et plus courants dans les rapports des petites organisations (57 %) que dans ceux des organisations de grande et de moyenne tailles (31 % et 26 % respectivement).

- Les constatations d'évaluation clés étaient résumées dans presque tous les résumés, soit complètement (50 %) soit partiellement (43 %).

- Les conclusions clés étaient également résumées dans la plupart des résumés, soit complètement (60 %) ou partiellement (26 %).

- Les recommandations d'évaluation étaient présentées complètement (69 %) ou partiellement (neuf pour cent) dans la majorité des résumés des rapports.

B) Introduction et contexte

Description

La grande majorité des évaluations des programmes de la fonction publique fédérale, soit 98 %, offrait une description claire et concise du programme, de la politique ou de l'initiative évalué (voir Tableau 1). Les notes de la qualité de la description de programme étaient fortes également : nous avons déterminé que 35 % des évaluations étaient adéquates selon ce critère et qu'une autre tranche de 49 % des évaluations offrait une discussion plus qu'adéquate.

La plupart des rapports décrivaient la totalité (64 %) ou une partie (29 %) des bénéficiaires visés et des intéressés du programme, de la politique ou de l'initiative. Nous avons déterminé que la plupart des rapports étaient adéquats (61 %) ou plus qu'adéquats (25 %) selon ce critère. Les rapports d'évaluation étaient un peu plus susceptibles d'avoir indiqué les bénéficiaires du programme (77 %) que ses intéressés (68 %).

La plupart des rapports d'évaluation de la fonction publique fédérale (71 %) incluaient une discussion de la répartition des ressources dans la description de programme. Nous avons déterminé que cette discussion était adéquate (37 %) ou plus qu'adéquate (40 %) dans ces rapports.

Environ six rapports d'évaluation de la fonction publique fédérale sur dix (59 %) décrivaient les hypothèses sous-jacentes du programme étudié (comme le financement, les partenariats) ou les facteurs externes (comme les influences environnementales). Dans les rapports (n=68) qui

déterminaient ces facteurs, 78 % décrivaient les hypothèses sous-jacentes du programme tandis que 66 % déterminaient les facteurs externes.

La principale faiblesse du volet de la description de programme était le manque de référence à un modèle de logique du programme : moins d'un rapport d'évaluation de la fonction publique fédérale sur quatre présentait un modèle de logique (19 % dans le rapport lui-même et un autre 4 % dans un document cité comme source de référence). À cet égard, seulement 22 % des rapports d'évaluation de la fonction publique fédérale incluaient une description des grands rapports de cause et d'effet sur lesquels le programme ou la politique était basé (par exemple, tel que présenté dans le modèle de logique). Dans les rapports qui incluaient une discussion de grands rapports de cause et d'effet (n=29), nous avons déterminé que la discussion était adéquate ou plus qu'adéquate dans la plupart des cas (41 % et 31 % respectivement), mais inadéquate dans le cas de 28 % des rapports.

TABLEAU 1 : Description du programme – Critères et Notes

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Décrire le programme, la politique ou l'initiative	98	16	35	49
Décrire les bénéficiaires et les intéressés	93*	14	61	25
Examiner l'attribution des ressources	71	23	37	40
Décrire les hypothèses sous-jacentes et les facteurs externes	59	10	59	30
Présenter un modèle logique	23**	s.o.	s.o.	s.o.
Décrire les grands rapports de cause et d'effet	22	28	41	31

Source : Examen des évaluations de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=29 à 113). « s.o. » signifie qu'il n'y a pas eu de note selon un critère. *La totalité ou une partie des bénéficiaires. **Présentation dans le rapport ou référence à un autre document.

Contexte d'évaluation

La grande majorité des rapports d'évaluation de la fonction publique fédérale (91 %) incluait une déclaration des objectifs de l'évaluation (Tableau 2). La notation de la qualité était élevée pour ce critère : nous avons déterminé que 52 % des rapports étaient adéquats et que 32 % étaient plus qu'adéquats à cet égard.

Environ six rapports sur dix (58 %) expliquaient le moment choisi pour l'évaluation (c'est-à-dire la période visée par l'étude) et une proportion semblable de rapports (56 %) décrivait la

signification de l'évaluation. Les rapports des ministères contenaient plus souvent une discussion de la signification de l'évaluation que les rapports des organismes (59 % contre 42 %) et les rapports des grandes organisations contenaient eux aussi plus souvent une discussion (65 %) que ceux des organisations de moyenne ou de petite taille (53 % et 39 % respectivement). La notation de la qualité pour ce critère était positive : nous avons déterminé qu'elle était plus qu'adéquate dans 30 % des cas, adéquate dans 59 % des cas et inadéquate dans 11 % des cas.

En ce qui a trait aux éléments et aux questions d'évaluation, les rapports d'évaluation de la fonction publique fédérale énumèrent simplement en général (les deux tiers) les questions (au lieu de les examiner, comme nous l'avons observé dans seulement 24 % des rapports). Cette pratique a fait en sorte que la note pour ce critère était comparativement faible par rapport aux autres notes. Pour ce critère, nous avons déterminé que 45 % des rapports étaient adéquats, que 20 % étaient plus qu'adéquats et que 35 % étaient inadéquats.

Une petite minorité des rapports d'évaluation de la fonction publique fédérale (8 %) a déterminé les éléments d'évaluation dans le contexte d'un Cadre de gestion et de responsabilisation axés sur les résultats (CGRR). Nous n'avons pratiquement pas noté de différences sur ce point selon le moment ou l'évaluation a été faite (avant ou après avril 2002). Toutefois, 42 % des rapports examinaient les éléments et les questions d'évaluation dans le contexte d'un autre document (habituellement un Cadre d'évaluation). La moitié des rapports n'ont pas fait référence à un contexte d'élaboration des éléments et des questions d'évaluation.

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Décrire les objectifs de l'évaluation	91	16	52	32
Décrire le moment choisi pour l'évaluation	58	s.o.	s.o.	s.o.
Décrire la signification de l'évaluation	56	11	59	30
Décrire les éléments et les questions d'évaluation	89*	35	45	20
Décrire le moment choisi pour l'évaluation	58	s.o.	s.o.	s.o.

Source : Examen des évaluations de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=64 à 106). « s.o. » signifie qu'il n'y a pas eu de note selon un critère. * Décrire ou énumérer les éléments.

En ce qui a trait à la couverture des éléments[7] (Tableau 3), la grande majorité des rapports d'évaluation de la fonction publique fédérale couvrait les éléments de réussite (94 %), suivis des éléments de pertinence (74 %) et des éléments de mise en œuvre et d'exécution (72 %). Les rapports sont beaucoup moins susceptibles d'avoir traité les pratiques de gestion (47 %) ou la rentabilité (44 %).

La couverture des éléments de la pertinence était plus courante dans les évaluations des petites et des moyennes organisations (89 % et 80 % respectivement) que dans les évaluations des grandes organisations (61 %). Les éléments de rentabilité étaient plus susceptibles d'être traités dans les évaluations faites après avril 2002 que dans celles faites avant (51 % par rapport à 27 %). Les évaluations des ministères étaient plus susceptibles que celles des organismes de traiter les éléments de pratiques de gestion (52 % par rapport à 29 %), comme les rapports des grandes et des moyennes organisations étaient plus susceptibles aussi de les traiter (50 % et 51 % respectivement) que ceux des petites organisations (28 %).

TABLEAU 3 : Couverture des éléments d'évaluation

Élément	Couverture (%)
Pertinence	74
Réussite	94
Rentabilité	44
Mise en œuvre et exécution	72
Pratiques de gestion	47

Source : Examen des évaluations de la fonction publique fédérale (n=115)

C) Méthodologie

Description de la méthodologie et de la conception

Les discussions de la méthodologie d'évaluation des rapports d'évaluation de la fonction publique fédérale étaient de qualité diverse – 56 % offraient une description complète de la méthodologie et de la conception de l'évaluation (Tableau 4). Quatre rapports sur dix énuméraient seulement quelques renseignements.

7 Outre les éléments d'évaluation de base du CT concernant la pertinence continue d'un programme, les résultats, les réussites et la rentabilité, certains rapports d'évaluation couvraient des éléments de la mise en œuvre et de l'exécution du programme (par exemple, la mesure selon laquelle les résultats attendus du programme étaient produits et offerts aux bénéficiaires visés) et les pratiques de gestion (par exemple, la pertinence de la structure de gouvernance du programme, la clarté des rôles de gestion, de responsabilités et de communications).

Dans la discussion de la méthodologie, les rapports étaient plus susceptibles d'établir la taille de l'échantillon (comme pour les entrevues avec les répondants clés et les enquêtes) (68 %). D'autre part, 45 % des rapports indiquaient la méthode d'échantillonnage, 30 % établissaient le lien entre les méthodes et les éléments et 26 % fournissaient des instruments de collecte des données. Le quart des rapports (27 %) ont fait référence à un document technique plus étoffé sur la méthodologie. Trois rapports sur dix ne contenaient rien de ce qui précède dans leur discussion méthodologique (c'est-à-dire la taille de l'échantillon, la méthode d'échantillonnage, les instruments, le lien entre les méthodes et les éléments, la référence aux documents techniques).

Le manque d'information méthodologique a affaibli la note de la qualité des rapports pour le critère en question : nous avons considéré que 46 % des rapports étaient inadéquats à cet égard, que 32 % étaient adéquats et que 21 % étaient plus qu'adéquats.

La moitié des rapports d'évaluation de la fonction publique fédérale (49 %) n'incluaient pas d'instruments de collecte des données et ne contenaient pas de référence à un document technique où trouver les instruments. Ce défaut était plus courant dans les évaluations des moyennes organisations (61 %) que dans celles des grandes ou des petites organisations (37 % et 44 % respectivement). Le quart des rapports (23 %) présentaient tous les instruments de recherche avec le rapport et quelques-uns des instruments étaient présentés dans une autre tranche de 10 % des rapports. Et 18 % des rapports faisaient référence à un document technique où il était possible de trouver les instruments.

Dans l'ensemble, la plupart des évaluations (72 %) étaient conçues convenablement en fonction des objectifs visés par l'étude (selon des considérations comme la rentabilité, la faisabilité et la validité). Il y avait 5 % des évaluations qui n'étaient pas conformes à ce critère et, dans 23 % des cas, l'examinateur a été incapable de faire une évaluation (à cause de la description inadéquate de la conception). Les évaluations considérées inadéquates avaient tendance à représenter seulement une gamme limitée de perspectives (par exemple, pas de participation du client, des entrevues avec des représentants de la fonction publique fédérale seulement) ou à avoir consulté seulement un petit nombre de particuliers et d'organisations.

Les notes de la qualité de la conception méthodologique étaient favorables : nous avons déterminé que 45 % des rapports d'évaluation étaient adéquats et que 42 % étaient plus qu'adéquats à cet égard. Nous avons déterminé que seulement 14 % de ces évaluations étaient inadéquates en ce qui a trait à la conception.

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Décrire les méthodologies et les conceptions appliquées	56	46	32	21
Éléments de description		s.o.	s.o.	s.o.
Taille de l'échantillon	68			
Méthode d'échantillonnage	45			
Liens entre les méthodes et les éléments	30			
Référence à des documents techniques	27			
Instruments	26			
Conception convenable	72	13	45	42

Source : Examen des évaluations de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=64 à 106). « s.o. » signifie qu'il n'y a pas eu de note selon un critère. * Décrire ou énumérer les éléments.

Sources de données multiples

Parmi les forces des évaluations de la fonction publique fédérale, pratiquement toutes les études (97 %) contenaient des sources multiples de données pour appuyer les constatations (Tableau 5). Nous avons déterminé que presque les deux tiers des rapports présentaient un équilibre adéquat entre les méthodologies qualitatives et quantitatives, tandis que 14 % présentaient un équilibre inadéquat (nous avons déterminé qu'environ les deux tiers de ces rapports dépendaient trop de méthodes qualitatives) et dans 23 % des cas, l'examinateur a été incapable de faire une évaluation.

Voici les méthodologies utilisées le plus fréquemment : entrevues avec des répondants clés (94 %), examens de documents (78 %), enquêtes sur échantillon (38 %), examens de dossiers (38 %), analyses documentaires (36 %), études de cas (35 %) et groupes de réflexion (24 %).

L'intégration de données d'un système de mesure continue du rendement n'était pas fréquente : 24 % des rapports indiquaient ces données comme sources de données pour l'évaluation.

Nous avons également déterminé que la majorité des rapports étaient de qualité adéquate (50 %) ou plus qu'adéquate (28 %) quant à l'inclusion d'une gamme de perspectives d'intéressés. Les évaluations des programmes de la fonction publique fédérale ont le plus souvent examiné la perspective de la gestion du programme et du personnel exécutant (83 %); les clients et les bénéficiaires (58 %); les partenaires (39 %); les bénéficiaires de financement (36 %); et les tiers

exécutants (24 %). Des experts ont également été consultés dans 20 % des évaluations; cette pratique était plus courante après avril 2002 qu'avant (24 % contre 11 %).

Seulement 26 % des rapports contenaient toutefois des données qualitatives obtenues de répondants clés qui n'avaient pas un intérêt dans le programme. Cette caractéristique méthodologique désirable était plus courante dans les évaluations faites après avril 2002 qu'avant (31 % contre 16 %) et dans les évaluations des petites et des moyennes organisations (39 % et 33 % respectivement) plutôt que dans celles des grandes organisations (13 %).

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Inclure des sources de données multiples	97	s.o.	s.o.	s.o.
Utiliser des données de surveillance continue du rendement	24	s.o.	s.o.	s.o.
Équilibrer convenablement l'aspect qualitatif et l'aspect quantitatif	64	s.o.	s.o.	s.o.
Inclure toutes les perspectives des intéressés*	s.o.	23	50	28
Inclure les perspectives des non intéressés	26	s.o.	s.o.	s.o.

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=115) « s.o. » signifie qu'il n'y a pas eu de note selon un critère. *Seuls les rapports pour lesquels ce critère pouvait être évalué ont été notés (n=97).

Limites

Quatre évaluations sur dix (39 %) des rapports d'évaluation de la fonction publique fédérale incluaient une discussion des limites des méthodologies et des sources de données utilisées (biais, fiabilité des données). Une proportion semblable de rapports (44 %) indiquaient les contraintes de l'évaluation, la disponibilité des données et le temps (34 % et 19 % respectivement) étant les contraintes mentionnées le plus souvent.

Rigueur

En ce qui a trait à la rigueur, peu d'évaluations des programmes de la fonction publique fédérale employaient la conception expérimentale traditionnelle ou quasi expérimentale. Si 44 % des évaluations incluaient un sondage représentatif des participants, seulement 13 % incluaient un groupe de comparaison et 14 % comparaient les données d'évaluation à une mesure de référence.

Une proportion un peu plus grande des rapports (22 %) incluait toutefois des données comparatives de la documentation ou un autre point de référence.

On note la tendance des évaluations des moyennes organisations à être un peu moins rigoureuses que celles des grandes ou des petites organisations. Par exemple, un sondage représentatif des participants et un groupe de comparaison étaient moins courants dans les évaluations des moyennes organisations (31 % et six pour cent respectivement) que dans celles des petites organisations (67 % et 22 %) ou dans celles des grandes organisations (50 % et 17 %).

D) Constatations clés

Pertinence

Juste un peu plus de la moitié des rapports d'évaluation (57 %) présentaient des données démontrant le besoin réel du programme en question et la réceptivité du programme à ce besoin (Tableau 6). Nous avons déterminé que la présentation de ces constatations était adéquate ou mieux pour 85 % des rapports examinés. La présentation de données sur ces deux éléments était moins courante dans les rapports des grandes organisations (46 % et 48 % respectivement) que dans ceux des moyennes organisations (61 % et 59 %) ou ceux des petites organisations (78 % pour les deux éléments). De plus, la qualité des données sur le second élément (réceptivité au besoin) était notée différemment selon la taille de l'organisation. Les rapports des petites et des grandes organisations étaient plus susceptibles d'être notés plus qu'adéquats à cet égard (47 % et 41 % respectivement) que les rapports des moyennes organisations (19 %). Notons également que ces éléments n'étaient simplement pas traités dans environ le tiers des évaluations.

TABLEAU 6 : Constatations de la pertinence – Critères et Notes

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Des données pour démontrer le besoin réel	57	15	45	40
Des données pour démontrer la réceptivité au besoin	57	13	54	32
Des données pour démontrer la pertinence continue par rapport aux priorités du gouvernement	58	12	47	41
Des données pour démontrer qu'il n'y a pas de double emploi	34	18	54	28

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=39 à 68).

De même, 58 % des rapports incluaient des données sur la pertinence continue du programme par rapport aux priorités du gouvernement et nous avons déterminé que la présentation de ces constatations était adéquate (47 %) ou plus qu'adéquate (41 %) pour la plupart des rapports. À nouveau, toutefois, la présentation de données sur cet élément de pertinence était moins courante dans les rapports des grandes organisations (48 %) que dans ceux des moyennes ou des petites organisations (environ les deux tiers dans chaque cas). Nous avons déterminé que moins de rapports des grandes organisations étaient plus qu'adéquats à cet égard (30 %) que ceux des petites ou des moyennes organisations (50 % et 46 % respectivement). Nous avons également déterminé que moins de rapports présentés avant avril 2002 étaient plus qu'adéquats que dans le cas des rapports présentés après cette date (32 % contre 46 %). Cet élément n'a pas été traité du tout dans 35 % des évaluations.

Concernant l'élément qui consiste à déterminer si le programme fait double emploi ou est en contraste avec d'autres programmes ou d'autres initiatives, seulement 34 % des évaluations présentaient des données et un groupe de 54 % n'a même pas traité cet élément. Pour les évaluations qui ont présenté certaines données, les notes étaient légèrement inférieures à celles des autres éléments de la pertinence : nous avons déterminé que 82 % des rapports étaient adéquats ou mieux, mais que 18 % étaient inadéquats à cet égard.

Réussite

La grande majorité des évaluations (87 %) a présenté des constatations démontrant si le programme, la politique ou l'initiative en question produisait ou non des résultats appuyant sa poursuite ou son renouvellement (Tableau 7). Seulement quatre pour cent des évaluations n'ont pas présenté ces constatations relatives à la réussite, et les éléments de réussite n'étaient pas traités dans le dernier groupe de neuf pour cent des évaluations. La proportion d'évaluations qui ont présenté des constatations relatives à la réussite était relativement plus élevée pour les petites organisations (100 %) que pour les moyennes et les grandes organisations (84 % et 85 % respectivement).

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Décrire les résultats du programme et l'attribution du programme par rapport à la réussite	87	26	37	37
Déterminer d'autres programmes, politiques et initiatives qui ont des liens, des résultats partagés	37	s.o.	s.o.	s.o.
Prendre ce qui précède en considération dans l'attribution	19	s.o.	s.o.	s.o.
Examiner d'autres facteurs qui contribuent aux résultats	61	14	50	36
Examiner les résultats fortuits	25	14	60	21
Traiter l'incrémentalité	26	26	48	27

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=29 à 100). « s.o. » signifie qu'il n'y a pas eu de note selon un critère.

Nous avons déterminé qu'environ le tiers (37 %) des évaluations avaient décrit les résultats plus qu'adéquatement, une proportion semblable (37 %) adéquatement et 26 % inadéquatement. La proportion indiquant que la présentation des constatations était inadéquate était considérablement inférieure pour les grandes organisations (18 %) par rapport aux petites et aux moyennes organisations (28 % et 33 % respectivement); et pour les évaluations produites après avril 2002 que pour les évaluations produites avant (19 % contre 39 %).

Un peu plus du tiers des évaluations (37 %) identifiaient d'autres programmes, politiques ou initiatives susceptibles de présenter des similarités, des rapports, des résultats partagés, ou des effets inter-programmes prévus, ou tout cela. Environ la moitié des évaluations (51 %) ne l'a pas fait. La proportion d'évaluations qui n'identifiaient pas d'autres programmes était considérablement plus élevée pour les organismes (62 %) que pour les ministères (49 %).

Environ le cinquième des évaluations (19 %) a pris d'autres programmes ou d'autres initiatives en considération dans la mesure de la réussite (attribution). Trois évaluations sur cinq (58 %) ne l'ont pas fait. La proportion des évaluations qui prennent d'autres programmes en considération augmente avec la taille de l'organisation, de 6 % pour les petites organisations, de 18 % pour les moyennes organisations et de 24 % pour les grandes organisations.

Trois évaluations sur cinq (61 %) examinaient d'autres facteurs qui contribuent aux résultats, tandis qu'environ le tiers des évaluations (31 %) ne l'a pas fait. Les petites organisations (72 %) étaient plus susceptibles de considérer d'autres facteurs contributifs que les organisations des autres tranches de taille (59 % pour les moyennes et les grandes organisations). De plus, les organismes étaient considérablement plus susceptibles de considérer d'autres facteurs que les ministères (75 % contre 57 %). Des proportions semblables identifiaient des facteurs internes et des facteurs externes.

Nous avons déterminé qu'environ le tiers (36 %) des évaluations avait considéré d'autres facteurs plus qu'adéquatement et que 50 % l'avaient fait adéquatement. Nous avons déterminé que seulement 14 % des évaluations avaient considéré les facteurs contributifs moins qu'adéquatement. La proportion des évaluations plus qu'adéquates était considérablement plus élevée pour les moyennes organisations (45 %) que pour les petites et les grandes organisations (31 % et 29 %).

Le quart des évaluations (25 %) ont considéré les résultats fortuits et environ les deux tiers (63 %) ne l'ont pas fait. Aucune différence importante n'a émergé des caractéristiques considérées. Parmi les évaluations qui mesuraient les résultats fortuits, environ la moitié a considéré les résultats positifs et environ la moitié a considéré les résultats négatifs.

Nous avons déterminé qu'environ les deux tiers des évaluations (66 %) examinaient adéquatement les résultats fortuits et que le cinquième (21 %) les examinait plus qu'adéquatement. Il y avait trop peu d'observations pour considérer les différences des résultats selon la taille et le type d'organisation et le moment choisi pour l'évaluation.

Le quart des évaluations (26 %) a considéré les résultats par incrémentalité tandis que presque les deux tiers (64 %) ne l'ont pas fait. La mesure de l'incrémentalité était considérablement plus élevée pour les organismes que pour les ministères (38 % contre 23 %) et pour les évaluations faites après avril 2002 que pour les évaluations faites avant (30 % contre 17 %). Parmi les évaluations qui ont évalué l'incrémentalité, 72 % ont examiné l'élément subjectivement et 28 % l'ont fait objectivement. Nous avons déterminé que l'incrémentalité était traitée adéquatement dans 53 % des rapports et plus qu'adéquatement dans 27 % des rapports. Il y avait trop peu d'observations pour considérer les différences dans les résultats selon la taille et le type d'organisation ou selon le moment choisi pour l'évaluation.

Rentabilité

Environ le quart des évaluations (26 %) examinait des démarches de remplacement qui pourraient produire des façons plus rentables de réaliser les résultats. Une proportion de 16 % des évaluations ne l'a pas fait et 58 % des évaluations n'ont pas traité la rentabilité. La proportion des évaluations qui a traité des démarches de remplacement baisse fortement selon la taille de l'organisation, dans une proportion de 50 % pour les petites organisations et dans une proportion

de 13 % pour les grandes organisations. De plus, cette proportion est beaucoup plus importante pour les évaluations faites après avril 2002 que pour les évaluations faites avant (31 % contre 16 %) et un peu plus importante pour les organismes que pour les ministères (38 % contre 23 %).

Parmi les évaluations qui ont considéré des démarches de remplacement rentables, nous avons jugé que 42 % les avaient évaluées adéquatement et que 29 % l'avaient fait plus qu'adéquatement. Encore une fois, il y avait trop peu d'observations pour considérer les différences dans les résultats selon la taille et le type d'organisation ou selon le moment choisi pour l'évaluation.

Parmi les évaluations qui ont considéré la rentabilité, environ deux fois plus d'évaluations l'ont considéré qualitativement que quantitativement. Ce ratio n'a pas varié beaucoup dans les caractéristiques en question, sauf qu'il était légèrement inférieur dans les grandes organisations. Nous avons déterminé qu'environ la moitié (49 %) des évaluations qualitatives ou quantitatives de la rentabilité des évaluations avait été exécutée adéquatement et que le quart (23 %) l'avait été plus qu'adéquatement. Nous avons déterminé cependant que 28 % de ces évaluations étaient inadéquates. Il y avait trop peu d'observations pour nous permettre d'observer à quel point la rentabilité a été traitée dans les caractéristiques des organisations.

Exécution et mise en œuvre

La plupart des évaluations (81 %) ont présenté des constatations relatives à la pertinence du modèle d'exécution, ou des pratiques de gestion, ou tout cela, pour contribuer aux objectifs du programme. Plus précisément, environ les deux tiers des évaluations (64 %) ont évalué le modèle d'exécution et 50 % ont examiné les pratiques de gestion. Une évaluation de ce dernier élément était plus courante dans les rapports des moyennes et des grandes organisations (55 % et 52 % respectivement) que dans les rapports des petites organisations (33 %). La présentation de ces constatations relatives à l'exécution et à la mise en oeuvre a obtenu une note élevée : nous avons déterminé que 50 % des évaluations étaient adéquates et que 39 % étaient plus qu'adéquates. Les évaluations plus qu'adéquates étaient beaucoup plus nombreuses pour les grandes que pour les petites organisations (43 % contre 29 %).

En outre, la plupart des évaluations (76 %) présentaient des données relatives au besoin d'améliorer les structures de programme ou les ententes d'exécution. Dans le cas de 14 % des évaluations examinées, les éléments d'exécution et de mise en œuvre n'ont pas été traités.

Autres aspects des constatations et de l'analyse

Dans la plupart des évaluations examinées, les éléments et les questions d'évaluation ont été traités adéquatement (47 %) ou plus qu'adéquatement (31 %), quoique 23 % des évaluations aient été inadéquates selon ce critère (Voir Tableau 8). De plus, en ce qui a trait à la présentation de constatations fondées sur des données qui découlent logiquement de données et d'analyses, la

majorité des évaluations ont été jugées adéquates ou mieux (46 % et 33 % respectivement), même si environ le cinquième (21 %) ont été jugées inadéquates. Les rapports des petites organisations étaient plus susceptibles d'être jugés plus qu'adéquats à cet égard (44 %) que ceux des grandes ou des moyennes organisations (36 % et 26 % respectivement). De plus, les évaluations faites après avril 2002 étaient un peu plus susceptibles d'être plus qu'adéquates selon ce critère que celles faites avant (37 % contre 24 %).

En ce qui a trait à la pertinence de l'analyse (c'est-à-dire la mesure dans laquelle l'analyse est appuyée par les données selon les tests de signification, les taux de réponse, etc.), les notes étaient relativement peu élevées. Premièrement, nous avons été incapables de faire cette évaluation pour 50 % des rapports d'évaluation, ce qui laisse croire que les renseignements clés relatifs à l'analyse ne sont pas inclus dans les rapports d'évaluation. Deuxièmement, parmi les rapports que nous avons évalués, environ les deux tiers étaient adéquats ou mieux (47 % et 21 % respectivement), mais 32 % étaient quand même inadéquats selon ce critère clé. Les raisons qui nous ont incités à considérer que l'analyse était inadéquate étaient notamment les suivantes : ne pas attribuer les constatations à des groupes distincts spécifiques qui avaient été consultés; ne pas indiquer la magnitude d'une constatation (par exemple, la proportion générale des intéressés qui pouvaient avoir une certaine opinion); trop se fonder sur l'analyse qualitative et anecdotique; et présenter des données d'échantillons de très petites tailles sans les accompagner des mises en garde qui s'imposent. Sur une note plus encourageante, nous avons déterminé qu'un moins grand nombre des évaluations faites après avril 2002 étaient inadéquates que dans le cas des évaluations faites avant (26 % contre 41 %), ce qui dénote une certaine amélioration. Les évaluations inadéquates étaient plus fréquentes pour les organismes (55 %) que pour les ministères (26 %).

Tableau 8 : Autres aspects des constatations et de l'analyse – Notes

Critères	Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Les éléments et les questions d'évaluation sont traités adéquatement	23	47	31
Les constatations sont fondées sur les données et découlent logiquement de l'interprétation des données et de l'analyse	21	46	33
L'analyse est convenable	32	47	21

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=57 à 115)

E) Conclusions clés

La plupart des évaluations présentaient des conclusions sur la pertinence (57 %) et la réussite (80 %) du programme ou de l'initiative en question, mais seulement 29 % ont tiré des conclusions sur la rentabilité. Il faut noter que moins d'évaluations des grandes organisations ont présenté des conclusions sur la pertinence ou la réussite (41 % et 70 % respectivement) que dans le cas des évaluations des petites organisations (67 % et 89 %) ou des moyennes organisations (67 % et 86 %). Parmi les évaluations qui ont tiré des conclusions sur ces trois éléments, nous avons déterminé que la plupart étaient adéquates (49 %) ou plus qu'adéquates (27 %) quant à la prestation de conclusions objectives, fondées sur des données, même si nous avons déterminé que 24 % des évaluations étaient inadéquates selon ce critère (Tableau 9). Nous avons déterminé qu'un nombre un peu plus élevé d'évaluations des grandes organisations étaient inadéquates (31 %) que d'évaluations des petites ou des moyennes organisations (environ le cinquième dans chacun des cas). Nous avons également déterminé que plus d'évaluations faites après avril 2002 étaient plus qu'adéquates selon ce critère que les évaluations faites auparavant (30 % contre 20 %), signe d'une certaine amélioration.

TABLEAU 9 : Conclusions – Critères et Notes

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Présenter des conclusions objectives, fondées sur des données, sur la pertinence, la réussite, ou la rentabilité, ou tout cela	s.o.	24	49	27
Présenter des conclusions objectives, fondées sur des données, sur la mise en œuvre et l'exécution, ou sur les pratiques de gestion, ou tout cela	s.o.	15	52	33
Présenter d'autres leçons apprises	54	5	54	41
Les conclusions sont fondées sur des critères de jugement ou des points de référence explicites	21	s.o.	s.o.	s.o.

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=56 à 96). « s.o. » signifie qu'il n'y a pas eu de note selon un critère.

Presque les deux tiers des évaluations ont tiré des conclusions sur la mise en œuvre et l'exécution (64 %), mais moins de la moitié a traité les pratiques de gestion dans les conclusions (44 %). Les conclusions relatives à ce dernier élément étaient moins courantes dans les évaluations des petites organisations (22 %) que dans celles des grandes ou des moyennes organisations (44 % et

53 % respectivement); dans les évaluations des organismes plutôt que dans celles des ministères (33 % contre 47 %); et dans les évaluations faites après avril 2002 que dans celles faites avant (40 % par rapport à 54 %). Les notes que nous avons déterminées pour la prestation de conclusions objectives, fondées sur des données, sur ces deux éléments étaient très fortes : nous avons déterminé que la majorité des évaluations étaient adéquates (52 %) ou plus qu'adéquates (33 %). Les notes élevées, plus qu'adéquates, étaient plus courantes pour les évaluations des grandes organisations (45 %) que pour celles des petites ou des moyennes organisations (environ le quart dans chaque cas) et pour les évaluations faites après avril 2002 (40 %) que pour les évaluations faites avant ce temps (20 %).

Environ la moitié des évaluations (49 %) présentait d'autres leçons apprises au sujet du programme. Dans le cas de ces rapports, les notes étaient très élevées quant à cet aspect. Nous avons déterminé que juste un peu plus de la moitié (54 %) des rapports étaient adéquats et que 41 % étaient plus qu'adéquats. Les notes les plus élevées (c'est-à-dire plus qu'adéquates) étaient plus courantes pour les évaluations faites après avril 2002 que pour celles faites avant (47 % contre 25 %).

Les conclusions de l'évaluation étaient clairement fondées sur des critères de jugement ou des points de référence explicites dans le cas de seulement une minorité (21 %) des évaluations, quoique nous ayons été incapables de faire une évaluation sur ce point pour 34 % des rapports (par exemple à cause du manque d'information). Nous avons observé l'absence de tels critères et de tels points de référence dans le cas de 45 % de l'ensemble des évaluations, et ce défaut était plus courant pour les évaluations faites avant avril 2002 que pour celles faites après (57 % contre 40 %).

F) Recommandations

Les trois quarts des rapports d'évaluation examinés contenaient des recommandations formelles (77 %). Une proportion additionnelle de 13 % des rapports proposait des mesures à prendre, sans indiquer toutefois qu'il s'agissait de recommandations. Seulement 10 % des rapports ne contenaient pas de recommandations ou de propositions. Les recommandations formelles étaient plus susceptibles d'apparaître dans les rapports des petites et des moyennes organisations (89 % et 86 % respectivement) que dans ceux des grandes organisations (63 %). Les rapports faits à compter d'avril 2002 étaient plus susceptibles de contenir des recommandations formelles que ceux faits avant (83 % contre 65 %). Finalement, les rapports produits par les organismes étaient plus susceptibles de contenir des recommandations formelles que ceux des ministères (88 % contre 75 %).

Parmi les rapports qui contenaient des recommandations (n=99), 26 % indiquaient des scénarios de remplacement et 35 % tenaient compte des contraintes pratiques comme les règlements et les budgets. Si nous avons déterminé que seulement 36 % des rapports étaient détaillés, nous avons également déterminé que les deux tiers étaient opérationnels (67 %) et que juste un peu moins

des deux tiers étaient pratiques (61 %). Les recommandations dans les rapports produits à compter d'avril 2002 étaient plus susceptibles d'être opérationnelles et pratiques que celles des rapports antérieurs (72 % contre 57 % et 65 % contre 51 % respectivement). Les recommandations dans les rapports des organismes étaient plus susceptibles d'être opérationnelles que celles dans les rapports des ministères (79 % contre 64 %).

Presque tous les rapports qui contenaient des recommandations (94 %) ont traité des constatations importantes (c'est-à-dire des constatations clés relatives aux grands éléments d'évaluation prioritaires), quoique neuf pour cent aient également traité des constatations qui n'étaient pas importantes. Nous avons déterminé de plus que la grande majorité des recommandations (94 %) découlaient logiquement des constatations et des conclusions de l'évaluation (Tableau 10).

Le quart des rapports qui contenaient des recommandations incluait une recommandation relative au financement global, et dans tous ces cas, la recommandation était d'accroître le financement. De plus, aucun rapport ne présentait de données indiquant qu'un programme n'était pas pertinent ou qu'il n'était pas nécessaire. Tous les rapports qui ont présenté des données sur les éléments de pertinence ont présenté des données indiquant que le programme évalué était pertinent et nécessaire. Il faut toutefois noter que ces constatations étaient parfois accompagnées de recommandations ou de propositions indiquant qu'il était nécessaire de restructurer ou de faire d'autres changements, mais toujours dans le contexte d'un programme encore pertinent et nécessaire.

TABLEAU 10 : Recommandations – Critères et Notes

Critères	Respect des critères (%)	Notes		
		Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Établir les scénarios de remplacement et tenir compte des contraintes pratiques	s.o.	35	48	17
Présenter des recommandations détaillées et opérationnelles (et pratiques)	s.o.	20	51	29
Présenter des recommandations qui traitent les constatations importantes	94	13	57	30
Présenter des recommandations qui découlent logiquement des constatations et des conclusions	94	15	53	32
Inclure une recommandation relative au financement global	25	s.o.	s.o.	s.o.

Source : Examen des évaluations des programmes de la fonction publique fédérale (n=115). Seuls les rapports qui respectaient les critères ont été notés (n=99 à 103). « s.o. » signifie qu'il n'y a pas eu de note selon un critère.

G) Réponse de la direction et Plan d'action

Juste un peu moins de la moitié des rapports d'évaluation examinés contenait une réponse de la direction (48 %). Les autres rapports (52 %) n'en contenaient pas.

Le tiers des rapports d'évaluation examinés contenait un plan d'action en réponse à l'évaluation (33 %). Les autres rapports (67 %) ne contenaient pas cet élément.

H) Clarté et autres aspects du rapport

Nous avons déterminé que les rapports d'évaluation étaient rédigés en général clairement, que 42 % étaient adéquats et que 44 % étaient plus qu'adéquats (Tableau 11). Notons que nous avons déterminé que 17 % des rapports étaient excellents à ce sujet. Une proportion de 22 % des rapports contenaient un lexique des acronymes pour les rendre plus clairs. Les rapports présentés à compter d'avril 2002 ont obtenu de meilleures notes que ceux qui ont été présentés avant cette date (53 % contre 24 % de rapports plus qu'adéquats).

En ce qui a trait à la présentation de l'information technique, nous avons déterminé que 55 % des rapports contenaient suffisamment mais non excessivement d'information dans le corps du rapport et que les annexes de 38 % des rapports contenaient de l'information technique pertinente et pratique (il faut noter que ces deux aspects ne sont pas mutuellement exclusifs). Nous avons toutefois déterminé que le tiers des rapports (33 %) étaient inadéquats quant à la pertinence de la présentation de l'information technique.

Nous avons déterminé que la grande majorité des rapports qui incluaient des annexes techniques (n=72) était de bonne qualité (69 % étaient adéquats et 18 % étaient plus qu'adéquats).

Une proportion de 43 % des rapports d'évaluation que nous avons examinés comptait entre 25 et 40 pages, une longueur jugée raisonnable aux fins de ces rapports. En ce qui a trait aux autres rapports, 20 % comptaient moins de 25 pages et 37 % en comptaient plus.

Tableau 11 – Clarté et autres aspects du rapport – Notes

Critères	Inadéquat (%)	Adéquat (%)	Plus qu'adéquat (%)
Rédiger le rapport d'évaluation avec clarté	15	42	44
Présenter l'information technique de façon convenable	30	51	18
Présenter des annexes techniques de qualité élevée	13	69	18
Présenter les données de façon acceptable	33	46	21
Utiliser efficacement les tableaux et les graphiques	25	52	23
Bien organiser le rapport pour qu'il soit facile à consulter	19	49	32

Source Examen des évaluations des programmes de la fonction publique fédérale (n=72 à 115)

Les rapports avaient tendance à être seulement modérément bien en ce qui a trait à la présentation des données. Nous avons déterminé que le tiers des rapports étaient inadéquats quant à la présentation acceptable des données (33 %) et que 25 % étaient pareillement inadéquats quant à l'utilisation efficace des tableaux et des graphiques. Nous avons déterminé en ce qui a trait à ces deux caractéristiques que juste un peu moins du quart des rapports étaient plus qu'adéquats. Nous avons cependant déterminé que la plus grande partie des rapports étaient adéquats quant à la présentation acceptable des données et à l'utilisation efficace des tableaux et des graphiques (46 % et 52 % respectivement). De plus, malgré les notes modérées qui précèdent, 65 % des rapports présentaient des chiffres et 71 % documentaient les sources de données.

Nous avons finalement déterminé, en ce qui a trait à la bonne organisation des rapports et à la facilité de les consulter, que presque le tiers étaient plus qu'adéquats (33 %) et que presque la moitié étaient adéquats (49 %). Les rapports présentés à compter d'avril 2002 étaient plus susceptibles d'être plus qu'adéquats en ce qui a trait à cette caractéristique que ceux qui ont été présentés avant cette date (39 % contre 16 %).

I) Évaluation globale

À la fin de chaque examen, l'examinateur a noté subjectivement la qualité globale du rapport d'évaluation. Nous avons déterminé que la plupart des rapports d'évaluation étaient adéquats (45 %) ou plus qu'adéquats (32 %), quoique seulement huit pour cent aient été « excellents ». Nous avons déterminé d'autre part que juste un peu moins du quart (23 %) étaient globalement inadéquats.

Il n'y avait pas de modèle clair des différences dans l'évaluation globale par rapport à la taille de l'organisation (par exemple, les rapports des petites organisations étaient à la fois plus susceptibles d'être inadéquats et d'être plus qu'adéquats que ceux des grandes organisations, qui étaient plus susceptibles d'être adéquats que ceux des petites organisations). Les rapports étaient toutefois plus susceptibles d'être inadéquats s'ils avaient été présentés avant avril 2002 (32 %, par rapport à 18 % pour avril 2002 et par la suite) et plus susceptibles d'être plus qu'adéquats s'ils avaient été présentés à compter d'avril 2002 et par la suite (37 % contre 22 % des rapports présentés avant cette date).

3.3 Forces et faiblesses des évaluations de la fonction publique fédérale

A) Forces

Voici les résumés des principales forces des évaluations de la fonction publique fédérale que nous avons déterminées :

- La plupart des rapports d'évaluation examinés présentaient bien le programme ou l'initiative évalué, y compris ses ressources, ses bénéficiaires et ses intéressés. Environ six rapports sur dix traitaient des hypothèses sous-jacentes du programme (comme le financement, les partenariats) et de facteurs externes comme les influences environnementales. La plupart des rapports incluaient aussi une déclaration claire des objectifs de l'évaluation.

- La majorité des évaluations (72 %) employaient une conception convenable de recherche, à la lumière des objectifs de l'étude, quoique nous ayons été incapables de faire une évaluation en fonction de ce critère pour presque le quart des rapports à cause du manque de renseignements. Parmi les rapports évalués, nous avons déterminé que la qualité de la conception de la méthodologie était adéquate ou mieux pour 87 % des évaluations. À peu près toutes les évaluations (97 %) incluaient des sources de données multiples.

- Plus de la moitié des évaluations (juste un peu moins de 60 %) présentaient des constatations relatives à la nécessité de poursuivre le programme et à sa pertinence. De ces évaluations, nous avons déterminé que la plupart (85 %) étaient adéquates ou plus qu'adéquates en fonction de ces critères.

- La majorité des évaluations (87 %) ont présenté des constatations démontrant si le programme ou l'initiative en question produisait ou non des résultats appuyant sa poursuite ou son renouvellement. Même si nous avons déterminé qu'environ le quart de ces rapports (26 %) étaient inadéquats en fonction de ce critère, la proportion des rapports dont la présentation de ces résultats était moins qu'adéquate a diminué (19 % après avril 2002 contre 39 % avant avril 2002).

- En ce qui a trait aux éléments d'exécution et de mise en œuvre, la plupart des évaluations ont présenté des constatations relatives à la pertinence du modèle d'exécution du programme, ou des pratiques de gestion, ou les deux (81 %), et à la nécessité d'améliorer la structure du programme ou ses modes d'exécution (76 %). Les évaluations étaient très bien notées en fonction du premier critère (89 % étaient adéquates ou plus qu'adéquates).

- Parmi les évaluations qui traitaient ces éléments, nous avons déterminé que la plupart (85 %) étaient adéquates ou mieux dans la prestation de conclusions objectives, fondées sur des données relatives à la mise en œuvre et à l'exécution, ou aux pratiques de gestion, ou tout cela. De plus, la qualité des évaluations s'améliore en ce qui a trait à ce critère : nous avons déterminé que 40 % des évaluations faites après avril 2002 étaient plus qu'adéquates à cet égard par rapport à seulement 20 % des rapports produits avant cette date.

- Dans leurs conclusions, la moitié des évaluations (49 %) présentaient d'autres leçons apprises au sujet du programme. Parmi ces rapports, nous avons déterminé que 95 % étaient adéquats ou plus qu'adéquats à cet égard.

- La grande majorité des évaluations incluaient des recommandations formelles (77 %) ou des suggestions de mesures à prendre (13 %). Dans presque tous les cas, les recommandations traitaient des constatations d'évaluation importantes (c'est-à-dire des constatations clés relatives aux grands éléments d'évaluation) et découlaient logiquement des constatations et des conclusions (94 % dans chaque cas).

- Nous avons déterminé que la plupart des rapports d'évaluation étaient adéquats ou plus qu'adéquats quant à la clarté de la rédaction (86 %) et à la bonne organisation (81 %).

B) Faiblesses

Voici les principales faiblesses ou les principaux aspects des évaluations de la fonction publique fédérale incluses dans le présent examen qui ont besoin d'amélioration :

- Les résumés ont besoin d'une certaine amélioration. Nous avons déterminé que le quart des résumés que nous avons examinés étaient inadéquats à titre de document cohérent et autonome et qu'environ le tiers ne présentaient pas du tout les éléments d'évaluation – quoique ce dernier défaut ait été moins courant dans les rapports présentés après avril 2002 (22 %) que dans ceux présentés avant (56 %).

- La plupart des rapports ne présentaient pas de modèle logique ou n'en indiquaient pas la référence et n'examinaient pas les grands rapports de cause et d'effet sur lesquels le programme était basé (moins du quart des évaluations incluaient ces éléments).

- Même si environ six rapports d'évaluation sur dix expliquaient le moment choisi pour l'évaluation et la signification de celle-ci, il semblerait qu'une proportion plus élevée de rapports devrait inclure de tels détails de base.

- La plupart des rapports (les deux tiers) ne faisaient qu'énumérer les éléments d'évaluation et très peu (environ le quart) en discutaient. De plus, la moitié des rapports ne faisaient référence à aucun document, comme un CGRR ou un Cadre d'évaluation, comme contexte d'élaboration des éléments d'évaluation.

- Moins de la moitié des rapports d'évaluation (44 %) traitaient les éléments de rentabilité, quoique la couverture de ces éléments ait été plus courante dans les évaluations faites après avril 2002 que dans celles faites avant (51 % contre 27 %).

- Un grand nombre de rapports ne contenaient pas de description complète des principaux aspects de la méthodologie. Si juste un peu plus de la moitié des rapports décrivaient la méthodologie, quatre sur dix en ont seulement énuméré quelques aspects. Seulement le quart des rapports ont fait référence à un document technique plus étoffé sur la méthodologie. Nous avons déterminé par conséquent que 46 % des rapports étaient inadéquats du point de vue de

leur description méthodologique. De plus, la moitié des rapports ne contenaient pas d'instruments de collecte des données ou une référence indiquant où trouver les instruments.

- Seulement une minorité des évaluations comprenaient des données d'un système de mesure du rendement (24 %) ou des données d'entrevue avec des répondants indépendants clés qui n'avaient pas un intérêt dans le programme (26 %). Cette dernière caractéristique est toutefois plus courante dans les évaluations faites après avril 2002 que dans celles faites avant (31 % contre 16 %).

- Malgré le fait que nous avons déterminé que presque les trois quarts des évaluations avaient une conception convenable de recherche pour les objectifs de l'étude, seulement une minorité des concepts d'évaluation incluaient des caractéristiques d'optimisation de la rigueur de l'analyse comme un groupe de comparaison (13 %), des mesures de référence (14 %) ou une comparaison avec des normes, de la documentation ou un autre point de référence (22 %).

- Seulement environ quatre rapports d'évaluation sur dix comprenaient une déclaration des limites ou des contraintes de l'évaluation.

- Seulement environ le tiers des évaluations présentaient des constatations permettant de déterminer si le programme faisait double emploi ou s'il était en contraste avec d'autres programmes ou d'autres initiatives.

- Seulement le quart des évaluations examinaient les résultats fortuits (25 %) ou traitaient les impacts différentiels (26 %). Environ les deux tiers des évaluations n'ont pas traité ces éléments.

- Seulement 26 % des évaluations présentaient des constatations concernant des démarches de remplacement susceptibles d'être plus rentables, quoique la couverture de cet élément ait augmenté dans les rapports les plus récents (31 % après avril 2002 contre 16 % avant avril 2002). De plus, environ le tiers des évaluations (34 %) présentaient une évaluation qualitative, ou quantitative, ou les deux, de la rentabilité du programme ou de l'initiative examiné, même si nous avons déterminé que 28 % de ces évaluations étaient inadéquates en fonction de ce critère.

- Il était difficile d'évaluer la pertinence de l'analyse (c'est-à-dire la mesure dans laquelle l'analyse était appuyée par les données selon les tests de signification, les taux de réponse, etc.) pour 50 % des évaluations en raison du manque de détails présentés dans les rapports. Parmi les rapports évalués en fonction de ce critère, nous avons déterminé que presque le tiers (32 %) étaient inadéquats. Cette dernière proportion a toutefois diminué au cours des dernières années (26 % après avril 2002 contre 41 % avant avril 2002).

- Nous avons déterminé que presque le quart des évaluations (24 %) étaient inadéquates dans leur prestation de conclusions objectives, fondées sur des données concernant la pertinence, la réussite ou la rentabilité, ou tout cela.

- Parmi les rapports qui contenaient des recommandations, seulement 26 % indiquaient des scénarios de remplacement et seulement 35 % tenaient compte des contraintes pratiques (par exemple les règlements, les budgets). Nous avons déterminé que plus du tiers de ces rapports (35 %) étaient inadéquats en fonction de ce critère.

- Moins de la moitié des rapports d'évaluation incluaient une réponse de la direction (48 %) ou un plan d'action (33 %).

- Plus du tiers des rapports (37 %) étaient très longs et comptaient plus de 40 pages.

- Nous avons déterminé qu'une proportion considérable des rapports étaient inadéquats du point de vue de la présentation acceptable des données, y compris les chiffres et les sources (33 %), la présentation convenable de l'information technique (30 %) et l'utilisation efficace des tableaux et des graphiques (25 %).

3.4 Variations de la qualité selon les caractéristiques organisationnelles et la date du rapport

A) Taille de l'organisation

Nous avons observé un certain nombre de différences intéressantes selon la taille de l'organisation. Il n'y avait toutefois pas de modèle constant dans les résultats selon la taille de l'organisation. Il n'est pas ressorti que les organisations d'une catégorie de taille particulière ou d'une autre présentaient de façon constante des évaluations de qualité supérieure à celles des organisations des autres catégories de taille. Les grandes différences selon la taille incluaient ce qui suit :

- Les grandes et les moyennes organisations (83 % et 92 % respectivement) étaient plus susceptibles d'inclure un résumé que les petites organisations (78 %).

- Les résumés qui ne présentaient pas les éléments d'évaluation étaient plus courants dans les rapports des petites organisations (57 %) que dans ceux des grandes et des moyennes organisations (31 % et 26 % respectivement).

- Une discussion de la signification de l'évaluation était plus courante dans les rapports des grandes organisations (65 %) que dans ceux des moyennes ou des petites organisations (53 % et 39 % respectivement).

- La couverture des éléments de la pertinence était plus courante dans les évaluations des petites et des moyennes organisations (89 % et 80 % respectivement) que dans celles des grandes organisations (61 %). Le traitement des éléments relatifs aux pratiques de gestion était plus courant dans les rapports des grandes et des moyennes organisations (50 % et 51 % respectivement) que dans ceux des petites organisations (28 %).

- L'absence d'instruments de collecte des données dans le rapport et de la référence à un document technique qui contiendrait ces instruments était plus courante dans les évaluations des moyennes organisations (61 %) que dans celles des grandes ou des petites organisations (37 % et 44 % respectivement).

- La présentation de données qualitatives obtenues de répondants clés qui n'avaient pas un intérêt dans le programme était plus courante dans les évaluations des petites et des moyennes organisations (39 % et 33 % respectivement) que dans celles des grandes organisations (13 %).

- Un sondage représentatif des participants et un groupe de comparaison étaient moins courants dans les évaluations des moyennes organisations (31 % et six pour cent respectivement) que dans celles des petites (67 % et 22 %) ou celles des grandes organisations (50 % et 17 %).

- Les rapports des moyennes organisations étaient beaucoup moins susceptibles d'être plus qu'adéquats dans la présentation de données sur la réceptivité au besoin (19 %) que ceux des petites ou des grandes organisations (47 % et 41 % respectivement).

- La présentation de données sur l'élément de pertinence continue était moins courante dans les rapports des grandes organisations (48 %) que dans ceux des moyennes ou des petites organisations (environ les deux tiers dans chaque cas). Il est ressorti que moins de rapports des grandes organisations étaient plus qu'adéquats à cet égard (30 %) que de rapports des petites ou des moyennes organisations (50 % et 46 % respectivement).

- La proportion des évaluations qui ont présenté des constatations relatives à la réussite était un peu plus élevée pour les petites organisations (100 %) que pour les moyennes et les grandes organisations (84 % et 85 % respectivement).

- La proportion des évaluations inadéquates quant à la présentation des constatations était considérablement plus faible pour les grandes organisations (18 %) que pour les petites et les moyennes organisations (28 % et 33 % respectivement).

- La proportion des évaluations qui prenaient d'autres programmes en considération en évaluant les incidences augmente avec la taille de l'organisation, de six pour cent pour les petites organisations, 18 % pour les moyennes, et 24 % pour les grandes organisations.

- Les petites organisations (72 %) étaient plus susceptibles de considérer d'autres facteurs cntributifs que les organisations des autres catégories de taille (59 % pour les moyennes et les grandes organisations).

- La proportion des évaluations plus qu'adéquates dans la prise en considération de facteurs contributifs dans la mesure de la réussite était considérablement plus élevée pour les moyennes organisations (45 %) que pour les petites et les grandes organisations (31 % et 29 %).

- La proportion des évaluations qui ont évalué des démarches de remplacement baisse fortement selon la taille de l'organisation, de 50 % pour les petites organisations à 13 % pour les grandes organisations.

- Une évaluation des pratiques de gestion était plus courante dans les rapports des moyennes et des grandes organisations (55 % et 52 % respectivement) que dans ceux des petites organisations (33 %). Les évaluations plus qu'adéquates étaient beaucoup plus nombreuses pour les grandes que pour les petites organisations (45 % contre 29 %).

- En ce qui concerne la présentation de constatations fondées sur des données qui découlent logiquement de données et d'analyses, plus de rapports des petites organisations étaient plus qu'adéquats (44 %) que de rapports des grandes ou des moyennes organisations (36 % et 26 % respectivement).

- Les conclusions sur les pratiques de gestion étaient moins courantes dans les évaluations des petites organisations (22 %) que dans celles des grandes ou des moyennes organisations (44 % et 53 % respectivement).

- Les évaluations plus qu'adéquates pour la prestation de conclusions fondées sur des données concernant les éléments d'exécution et de pratiques de gestion étaient plus courantes dans le cas des grandes organisations (45 %) que dans le cas des petites ou des moyennes organisations (environ le quart dans chaque cas).

- Les recommandations formelles étaient plus susceptibles d'apparaître dans les rapports des petites et des moyennes organisations (89 % et 86 % respectivement) que dans ceux des grandes organisations (63 %).

B) Avant et après avril 2002

Nous avons noté des différences clés selon le moment où le rapport a été produit. En général, les évaluations faites après avril 2002 avaient des notes plus élevées que celles faites avant. Voici les résultats détaillés :

- Les résumés ne présentant pas les éléments d'évaluation étaient plus courants dans les rapports faits avant avril 2002 que dans ceux faits après (56 % contre 22 %).

- Les éléments de rentabilité étaient plus susceptibles d'être traités dans les évaluations faites après avril 2002 que dans celles faites avant (51 % contre 27 %).

- La présentation de données qualitatives obtenues de répondants clés qui n'avaient pas un intérêt dans le programme était plus courante dans les évaluations faites après avril 2002 que dans celles faites avant (31 % contre 16 %).

- Moins de rapports présentés avant avril 2002 étaient plus qu'adéquats en ce qui a trait aux constatations portant sur la pertinence continue, que dans le cas des rapports présentés après cette date (32 % contre 46 %).

- La proportion pour laquelle la présentation de constatations sur la réussite était inadéquate était considérablement moins élevée dans les rapports produits après avril 2002 que dans ceux produits avant (19 % contre 39 %).

- La proportion des évaluations qui traitaient les démarches de remplacement était beaucoup plus grande dans les évaluations faites après avril 2002 que dans celles faites avant (31 % contre 16 %).

- Les évaluations faites après avril 2002 étaient un peu plus susceptibles d'être plus qu'adéquates concernant la présentation de constatations fondées sur des données qui découlent logiquement de données et d'analyses que dans le cas des évaluations faites avant cette date (37 % contre 24 %).

- Moins d'évaluations faites après avril 2002 étaient inadéquates concernant la pertinence de l'analyse que dans le cas des évaluations faites avant cette date (26 % contre 41 %).

- Plus d'évaluations faites après avril 2002 étaient plus qu'adéquates concernant la prestation de conclusions objectives fondées sur des données (sur la pertinence, la réussite ou la rentabilité, ou tout cela) que pour les évaluations faites avant cette date (30 % contre 20 %), signe d'une certaine amélioration.

- Les conclusions sur les pratiques de gestion étaient moins courantes dans les rapports d'évaluation produits après avril 2002 que dans ceux produits avant cette date (40 % contre 54 %). Les rapports d'évaluation plus qu'adéquats concernant les conclusions sur les éléments d'exécution et de mise en œuvre étaient plus courants pour les évaluations faites après avril 2002 (40 %) que pour celles faites avant (20 %).

- Les rapports faits à compter d'avril 2002 étaient plus susceptibles de contenir des recommandations formelles que ceux faits avant (83 % contre 65 %).

- Les rapports présentés à compter d'avril 2002 étaient plus susceptibles que les rapports présentés avant cette date d'être plus qu'adéquats concernant la clarté de la rédaction (53 % contre 24 %).

- Les rapports étaient plus susceptibles d'être inadéquats dans l'ensemble s'ils avaient été présentés avant avril 2002 (32 % contre 18 % pour avril 2002 et par la suite) et plus susceptibles d'être plus qu'adéquats s'ils avaient été présentés à compter d'avril 2002 (37 % contre 22 % des rapports présentés avant cette date).

C) Organisme par rapport au ministère

Nous avons observé quelques différences entre les évaluations parrainées par les organismes et celles parrainées par les ministères, mais il n'y avait pas de modèle constant dans les résultats. Les différences entre les rapports des organismes et ceux des ministères étaient notamment les suivantes :

- La discussion de la signification de l'évaluation était plus courante dans les rapports des ministères que dans ceux des organismes (59 % contre 42 %).

- Le traitement des éléments relatifs aux pratiques de gestion était plus courant dans les évaluations des ministères que dans celles des organismes (52 % contre 29 %).

- La présentation de données obtenues de répondants clés qui n'avaient pas un intérêt dans le programme était plus courante dans les évaluations faites après avril 2002 que dans celles faites avant (31 % contre 16 %).

- Les évaluations des organismes étaient considérablement plus susceptibles de considérer d'autres facteurs ayant contribué aux résultats que celles des ministères (75 % contre 57 %).

- La mesure de l'incrémentalité était incluse dans plus d'évaluations des organismes que des ministères (38 % contre 23 %).

- Les évaluations des organismes étaient plus susceptibles de traiter des démarches de remplacement que celles des ministères (38 % contre 23 %).

- La présentation de conclusions concernant la mise en œuvre et l'exécution était moins courante dans les évaluations parrainées par les organismes que dans celles parrainées par les ministères (33 % contre 47 %).

- Les rapports produits pour les organismes étaient plus susceptibles de contenir des recommandations formelles que ceux produits pour les ministères (88 % contre 75 %).

- Les recommandations dans les rapports pour les organismes étaient plus susceptibles d'être opérationnelles que celles dans les rapports pour les ministères (79 % contre 64 %).

4. Conclusions et Recommandations

4.1 Conclusions

Dans l'ensemble, la plupart des évaluations que nous avons examinées sont de qualité raisonnable. La majorité a reçu la note globale adéquate (45 %) ou « plus qu'adéquate » (32 %). Mais une proportion considérable des évaluations (23 %) a reçu la note inadéquate et il s'agit d'une constatation sur laquelle le CEE doit se pencher. Nous n'avons pas observé de modèles clairs et constants quand nous avons comparé les rapports des organisations de tailles différentes ou les rapports des ministères par rapport à ceux des organismes. Nous avons toutefois observé une amélioration visible concernant un certain nombre de critères quand nous avons comparé les évaluations faites avant avril 2002 à celles faites après cette date. Les évaluations plus récentes démontrent une amélioration importante de la qualité, ce qui laisse croire que la Politique d'évaluation d'avril 2001 du SCT a pu avoir un effet favorable.

Comme nous l'avons illustré de façon détaillée dans le chapitre précédent, nous avons déterminé un certain nombre de forces dans les évaluations des programmes de la fonction publique fédérale. Les principales forces incluent : une description complète du programme ou de l'initiative examiné, y compris ses ressources, ses bénéficiaires et ses intéressés; une déclaration claire des objectifs d'évaluation; le recours à des sources multiples de données dans la méthodologie; une bonne présentation des constatations, en particulier, sur les éléments de la pertinence, de l'exécution et de la mise en œuvre; l'inclusion de recommandations formelles ou de propositions d'améliorations, avec des recommandations découlant logiquement des constatations et des conclusions; et des rapports bien rédigés et bien organisés.

D'autre part, notre examen a également révélé un certain nombre de faiblesses des évaluations et des rapports, incluant ce qui suit : négliger de présenter le modèle logique du programme ou ne pas en indiquer la référence; présenter une discussion inadéquate des éléments d'évaluation et ne pas indiquer la référence aux documents sources comme les CGRR ou les Cadres d'évaluation; décrire de façon inadéquate les aspects de la méthodologie et négliger de joindre les instruments de collecte des données ou d'en indiquer la référence; utiliser de façon inadéquate les données de surveillance du rendement et les points de vue des répondants clés indépendants qui n'ont pas un intérêt dans le programme; évaluer de façon inadéquate l'impact différentiel des programmes et ne pas utiliser suffisamment de groupes de comparaison et de mesures de référence dans les conceptions d'évaluation; couvrir superficiellement les éléments de rentabilité.

4.2 Recommandations

Sur la base des constatations du présent examen, voici ce qui est recommandé au CEE :

❶ Encourager les divisions des évaluations des ministères et des organismes fédéraux à renforcer leurs rapports d'évaluation en traitant les faiblesses majeures déterminées dans le présent examen.

Améliorer les rapports d'évaluation

- assurer que le résumé d'un rapport comprend tous les points clés et sert de résumé autonome des objectifs d'évaluation, des éléments, de la démarche méthodologique, des constatations clés, des conclusions et (s'il y a lieu) des recommandations;
- présenter le modèle de logique du programme dans le rapport ou dans une annexe, ou indiquer la référence pour le trouver (par exemple un CGRR, un Cadre d'évaluation);
- énumérer tous les éléments d'évaluation dans le rapport ou dans une annexe, ou fournir une référence pour la liste au complet;
- présenter tous les détails clés de la méthodologie (par exemple les méthodes utilisées, le moment choisi pour la collecte des données, le nombre de répondants, les types d'analyse) et les instruments de collecte des données, dans le rapport et ses annexes ou dans un document technique indiqué en référence;
- indiquer les limites de l'évaluation et ses contraintes;
- présenter les constatations et les données de façon acceptable en incluant les détails clés sur les données et l'analyse dans le rapport ou dans les annexes, en particulier, les taux de réponse, les tests de signification, les chiffres et les résultats quantitatifs et les sources de données;
- présenter des conclusions objectives fondées sur des données, qui sont liées clairement et logiquement aux constatations de l'évaluation sur laquelle elles sont fondées;
- dans les recommandations, considérer des scénarios de remplacement (s'il y a lieu) et les contraintes pratiques des mesures à prendre qui sont proposées;
- essayer de garder le corps du rapport d'évaluation d'une longueur de 25 à 40 pages et présenter l'information supplémentaire essentielle (par exemple les constatations détaillées et les analyses techniques) dans les annexes.

Améliorer les rapports d'évaluation

- consulter des répondants clés indépendants (qui n'ont pas un intérêt dans le programme) dans plus d'évaluations;
- incorporer une analyse des données de surveillance du rendement dans plus d'évaluations;
- incorporer des mesures de référence et un groupe de comparaison dans la conception de recherche des évaluations lorsque les impacts différentiels d'un programme sont un élément important; et
- inclure une évaluation quantitative des éléments de rentabilité dans plus d'évaluations finales et sommatives.

❷ Peaufiner les lignes directrices et les critères du Conseil du Trésor concernant les caractéristiques attendues des (1) méthodes d'évaluation et (2) des rapports d'évaluation, et les diffuser.

❸ Poursuivre la mise en oeuvre d'une démarche rigoureuse de surveillance de la qualité des évaluations et l'utiliser comme base d'élaboration de fiches de rendement individuelles sur la qualité et la santé globale de la fonction d'évaluation par ministère et par petit organisme.

❹ Indiquer des mesures, y compris une structure de stimulants et des normes, afin d'assurer que les ministères et les organismes présentent des évaluations et des examens d'une façon responsable et raisonnable. Le respect de ces normes par les ministères et les organismes devrait être annoncé publiquement.

Annexe A
Gabarit d'examen

Examen de la qualité des évaluations

Gabarit d'examen (Version finale : 7e ébauche)

Description du rapport d'évaluation

Numéro du rapport	
Ministère	Petit ❑ Moyen ❑ Grand ❑
Organisme	Petit ❑ Moyen/Grand ❑
Taille du Groupe d'évaluation de l'org.	
Type de rapport	❑ Examen
	❑ Évaluation formative
	❑ Évaluation sommative
	❑ Étude spéciale (p. ex., recherche)
	❑ Autre : _____
Date du rapport	
Examinateur	

Examen de la qualité des évaluations
Gabarit d'examen (26 avril 2004)

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
1.0 Résumé (Remarque : Évaluation dernière)			☐ Oui ☐ Non				
	1.1 Présenter une évaluation claire et concise, cohérente à titre de document autonome				Médiocre 1 2 Adéquate 3 4 Excellente 5		
	1.2 Présenter les éléments d'évaluation clés et répondre à ces éléments avec de l'information pertinente au moyen d'une analyse solide	Les **éléments** d'évaluation clés sont résumés	☐ Oui – complètement ☐ Oui – partiellement ☐ Non				
		Les **constatations** d'évaluation clés sont résumées	☐ Oui – complètement ☐ Oui – partiellement ☐ Non				
		Les **conclusions** d'évaluation clés sont résumées	☐ Oui – complètement ☐ Oui – partiellement ☐ Non				
		Les **recommandations** d'évaluation sont présentées	☐ Oui – complètement ☐ Oui – partiellement ☐ Non ☐ S.O.				

8 Une note de 3 indique que le critère est respecté, tandis qu'une note de 1 ou 2 indique que le critère n'est pas respecté convenablement. Une note de 4 ou 5 indique une qualité excellente qui signifie que les considérations minimales de base du critère sont dépassées ou extrêmement bien appliqués.

9 Évaluation qualitative à faire seulement lorsque ✓ apparaît dans la cellule.

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
2.0 Introduction et Contexte							
2.1 Description	2.1.1 Décrire clairement et avec concision le programme, la politique ou l'initiative évalué		☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	2.1.2 Décrire les bénéficiaires visés et les intéressés concernés		☐ Oui – tous ☐ Oui – quelques-uns ☐ Non	☐ bénéficiaires ☐ intéressés	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	2.1.3 Décrire les liens de cause et d'effet entre les intrants, les activités, les extrants et les résultats, et les facteurs externes contribuant à la réussite ou à l'échec	Présenter un modèle de logique dans le rapport	☐ Oui ☐ Non – mais référence fournie ☐ Non – pas de référence				
		Décrire les grands rapports de cause et d'effet (p.ex., comme présentés dans le modèle de logique)	☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Décrire les hypothèses sous-jacentes(p.ex., le financement, les partenariats) ou les facteurs externes (comme les influences environnementales) ou tout cela	☐ Oui ☐ Non	☐ hypothèses sous-jacentes ☐ facteurs externes	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
	2.1.4 Examiner l'attribution des ressources à la politique, au programme ou à l'initiative	Décrire clairement les ressources du programme pour faire comprendre comment les crédits du programme ont été attribués et dépensés	☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
2.2 Contexte d'évaluation	2.2.1 Déterminer le rôle de l'évaluation et son importance/sa signification au moment où elle a été exécutée	Décrire les objectifs de l'évaluation	☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Décrire le moment choisi pour l'évaluation	☐ Oui ☐ Non				
		Décrire la signification de l'évaluation	☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	2.2.2 Décrire les éléments d'évaluation clés et les questions liés au programme, à la politique ou à l'initiative	Décrire les éléments d'évaluation et les questions	☐ Oui – les éléments sont examinés ☐ Oui – les éléments sont seulement énumérés ☐ Non	☐ présenter les éléments dans une annexe technique	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
		Déterminer les éléments d'évaluation dans le contexte d'un CGRR ou d'autres documents clés	☐ Oui – CGRR ☐ Oui – autres documents ☐ Non ☐ Impossibilité d'évaluer				
		Couverture : › pertinence › réussite › rentabilité		☐ pertinence ☐ réussite ☐ rentabilité			
		Inclure les éléments concernant : › la mise en œuvre et l'exécution › les pratiques de gestion		☐ la mise en œuvre et l'exécution ☐ les pratiques de gestion			
3.0 Méthodologie							
3.1 Description de la méthodologie et de la conception	3.1.1 Décrire les méthodologies logiques, valides, fondées sur des données qui sont liées aux éléments d'évaluation explorés OU inclure une référence claire à un document technique contenant cette information	Décrire les méthodologies et la conception qui s'appliquent à l'évaluation	☐ Oui – description ☐ Oui – seulement l'énumération de quelques détails ☐ Non – pas de référence à des documents techniques ☐ Non – référence à des documents techniques	☐ taille de l'échantillon ☐ Méthode d'échantillonnage ☐ instruments ☐ liens entre les méthodes et les éléments ☐ référence à des documents techniques	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
		Présenter les instruments	☐ Oui – tous ☐ Oui – quelques-uns ☐ Non – pas de référence à des documents techniques ☐ Non – référence à des documents techniques				
		Utiliser une conception convenable pour les objectifs visés de l'étude (p.ex., rentable, faisable, logique, valide)	☐ Oui ☐ Non ☐ Impossibilité d'évaluer		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9	✓	
3.2 Sources de données multiples	3.2.1 Inclure des sources multiples de données dans l'évaluation pour appuyer la validité des constatations	Faire dépendre l'évaluation de plus d'une source de données pour appuyer ses constatations › qualitative › quantitative › analyse documentaire › examen des documents › examen des dossiers › analyse des données secondaires › examen des bases de données › analyse des données sur le rendement › études de cas › analyse de rentabilité › autre	☐ Oui ☐ Non – non mais elle aurait dû ☐ Non – mais ce n'est pas nécessaire ou convenable pour l'évaluation en question	☐ qualitative ☐ groupe de discussion ☐ entrevues avec des répondants clés ☐ autre _____ ☐ quantitative ☐ recensement ☐ taille de l'échantillon ☐ autre _____ ☐ analyse documentaire ☐ examen des documents ☐ examen des dossiers ☐ analyse des			

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
				données secondaires ☐ examen des bases de données ☐ analyse des données sur le rendement ☐ études de cas ☐ analyse de rentabilité ☐ autre _____			
		L'évaluation utilise des données d'un système de surveillance continue du rendement	☐ Oui ☐ Non – données existantes mais non utilisées ☐ Non – pas de données ☐ Sans objet ☐ Impossibilité d'évaluer				
	3.2.2 Est-ce que l'équilibre est convenable entre les méthodologies qualitatives et quantitatives?		☐ Oui ☐ Non ☐ S.O.			✓	
	3.2.3 Inclure les perspectives de tous les intéressés	› Clients/bénéficiaires › gestion et exécution du programme (fonction publique fédérale) › tiers exécutants › partenaires › experts › bénéficiaires de financement › non-bénéficiaires	☐ Impossibilité d'évaluer	☐ clients/ bénéficiaires ☐ gestion et exécution du programme (fonction publique fédérale) ☐ tiers exécutants ☐ partenaires ☐ experts	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
		› autre _____		☐ bénéficiaires de financement ☐ non-bénéficiaires ☐ autre _____			
		obtention de données qualitatives de répondants clés qui n'ont pas un intérêt dans le programme	☐ Oui ☐ Non ☐ Impossibilité d'évaluer				
3.4 Limites	3.4.1 Formuler clairement les limites et les compromis des méthodologies, des sources de données et des données utilisées dans l'évaluation	description des limites : détermination et explication des biais réels et possibles et la fiabilité des données du point de vue de leur incidence sur les constatations présentées	☐ Oui ☐ Non ☐ Pas de limites apparentes	☐ description des biais ☐ explication de la qualité des données	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Les contraintes de l'évaluation sont présentées clairement	☐ Oui ☐ Non ☐ Pas de contraintes apparentes	☐ budget ☐ temps ☐ disponibilité des données ☐ autre _____	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
3.5 Rigueur	3.5.1 Existence d'un « point » de comparaison	Enquête auprès d'un groupe représentatif de participants	☐ Oui ☐ Non				
		Groupe de comparaison	☐ Oui ☐ Non				
		Comparaison avec des mesures de référence	☐ Oui ☐ Non				
		Comparaison avec des normes, de la documentation, d'autres points de repère	☐ Oui ☐ Non				

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
4.0 Constatations clés							
4.1 Pertinence	4.1.1 Présenter des constatations concernant l'établissement de la pertinence continue et de la contribution à la réalisation des résultats en liant les résultats aux besoins sociétaux et aux priorités du gouvernement	Des données pour démontrer le besoin réel	☐ Oui ☐ Non ☐ Non traité		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Des données pour démontrer la réceptivité au besoin	☐ Oui ☐ Non ☐ Non traité		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Des données pour démontrer la pertinence continue par rapport aux priorités du gouvernement	☐ Oui ☐ Non ☐ Non traité		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Des données pour démontrer que le programme ne fait pas double emploi ou ne contraste pas avec d'autres programmes, politiques ou initiatives	☐ Oui ☐ Non ☐ Non traité		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
4.2 Réussite	4.2.1 Présenter des constatations qui démontrent si le programme, la politique ou l'initiative en question produit ou non des résultats qui appuient sa poursuite ou son renouvellement	Décrire clairement ce qui s'est produit comme résultat du programme et formuler la contribution du programme, de la politique ou de l'initiative à la réussite	☐ Oui		Médiocre 1		
	4.2.2 Déterminer d'autres programmes, politiques ou initiatives susceptibles de présenter des similarités, des rapports, des résultats partagés, ou des effets inter-programmes prévus, ou tout cela	Déterminer d'autres programmes, politiques ou initiatives	☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités				
		Prendre ce qui précède en considération dans l'attribution	☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités				
	4.2.3 Examiner d'autres facteurs qui contribuent aux résultats (p.ex. concernant le financement ou le partenariat, les facteurs externes)		☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités	☐ facteurs internes du programme ☐ facteurs externes	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	4.2.4 Examiner s'il s'est produit des résultats fortuits qui ont contribué à la réussite ou qui ont présenté des contraintes spécifiques		☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités	☐ résultats positifs ☐ résultats négatifs	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
	4.2.5 L'incrémentalité est traitée		☐ Oui	☐ subjectivement	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
4.3 Rentabilité	4.3.1 Déterminer la mesure dans laquelle le programme, la politique ou l'initiative aurait pu être exécuté avec des méthodes plus convenables et rentables pour réaliser ses objectifs	Examiner des démarches de remplacement qui pourraient produire des façons plus rentables d'obtenir les résultats	☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Présenter : › une évaluation qualitative de la rentabilité › une évaluation quantitative de la rentabilité	☐ Oui ☐ Non ☐ S.O. – les éléments de réussite ne sont pas traités	☐ évaluation qualitative ☐ évaluation quantitative	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
4.4 Exécution et mise en oeuvre	4.4.1 Présenter des constatations relatives à la détermination de l'efficacité et de la pertinence du champ des structures du programme et des ententes de prestation du service pour le programme, la politique ou l'initiative	Évaluer le modèle d'exécution, sa pertinence et sa contribution à la réalisation des objectifs › pratiques de gestion	☐ Oui ☐ Non ☐ S.O.	☐ modèle d'exécution ☐ pratiques de gestion	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Présenter des données permettant de déterminer s'il y a lieu d'améliorer les structures ou les ententes d'exécution du programme	☐ Oui ☐ Non ☐ S.O.				

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[a]	Évaluation qualitative[a]	Autres observations
4.5 Éléments d'évaluation	4.5.1 Traiter convenablement les éléments et les questions d'évaluation				Médiocre 1 2 Adéquate 3 4 Excellente 5		
4.6 Constatations fondées sur des données	4.6.1 Présenter des constatations fondées sur des données extraites de l'étude d'évaluation	Démontrer que les constatations découlent logiquement de l'interprétation des données et des analyses			Médiocre 1 2 Adéquate 3 4 Excellente 5		
4.7 Analyse	4.7.1 L'analyse est convenable	Les données soutiennent l'analyse (selon, par exemple, les tests de signification et les taux de réponse)	☐ Impossibilité d'évaluer		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9	✓	
5.0 Conclusions clés							
	5.1 Présenter des conclusions claires, impartiales et exactes fondées sur des données	Les conclusions répondent objectivement aux éléments d'évaluation et sont appuyées par les constatations		☐ pertinence ☐ réussite ☐ rentabilité	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
				☐ la mise en œuvre et l'exécution ☐ pratiques de gestion	Médiocre 1 2 Adéquate 3 4 Excellente 5		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
					S.O. 9		
		Présenter d'autres leçons apprises au sujet du programme à la lumière de l'évaluation	☐ Oui ☐ Non ☐ Impossibilité d'évaluer		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Les conclusions sont fondées sur des critères de jugement ou des points de repère explicites	☐ Oui ☐ Non ☐ Impossibilité d'évaluer	☐ pas de critères présentés			
6.0 Recommandations			☐ Oui – recommandations formelles ☐ Oui – des propositions qui ne sont pas appelées des « recommandations » ☐ Non				
	6.1 Présenter de façon claire des recommandations pratiques qui peuvent être réalisées	Établir les scénarios de remplacement et tenir compte des contraintes pratiques (p. ex., les règles, les établissements et le budget)		☐ scénarios de remplacement ☐ contraintes pratiques	Médiocre 1 2 Adéquate 3 4 Excellente 5		
		Présenter des recommandations détaillées et opérationnelles (et pratiques)		☐ détaillées ☐ opérationnelles ☐ pratiques	Médiocre 1 2 Adéquate 3 4 Excellente 5		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
	6.2 Présenter des recommandations qui découlent logiquement des constatations et des conclusions et sur lesquelles elles sont appuyées	Présenter des recommandations qui traitent les constatations importantes	☐ Oui	☐ traiter aussi les recommandations moins importantes	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
		Présenter des recommandations qui découlent logiquement des constatations et des conclusions	☐ Oui ☐ Non		Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	6.3 Inclure une recommandation relative au financement global		☐ Oui ☐ Non	☐ augmenter le financement ☐ diminuer le financement			
7.0 Réponse de la direction			☐ Oui ☐ Non				
8.0 Plan d'action			☐ Oui ☐ Non				
9.0 Aspects généraux/Autre							
9.1 Clarté	9.1.1 Le rapport est rédigé en clair et les annexes techniques fournissent de l'information technique détaillée	Rédiger le rapport d'évaluation avec clarté		☐ lexique des acronymes	Médiocre 1 2 Adéquate 3 4 Excellente 5		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
		Présenter l'information technique de façon convenable		☐ suffisamment d'information technique dans le corps du rapport mais non excessivement	Médiocre 1		
9.2 Autres aspects du rapport	9.2.1 Le corps du rapport est d'une longueur raisonnable (25 à 40 pages)		☐ Oui ☐ Non	☐ moins de 25 pages ☐ 25 à 40 pages ☐ plus de 40 pages			
	9.2.2 Les annexes techniques sont déterminées clairement, comme les endroits où elles se trouvent		☐ Oui – clairement ☐ Oui – mais pas assez clairement ☐ Non				
	9.2.3 Présenter des annexes techniques de qualité élevée				Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	9.2.4 Présenter les données de façon acceptable	Présenter les chiffres Documenter les sources		☐ présentation des chiffres ☐ documentation des sources	Médiocre 1 2 Adéquate 3 4 Excellente 5		

Éléments/Exigences	Critères	Considérations	Liste de vérification générale	Liste de vérification détaillée	Note[8]	Évaluation qualitative[9]	Autres observations
	9.2.5 Utiliser efficacement les tableaux et les graphiques	Bonne présentation Facile à lire acceptable	☐ Pas de tableaux ☐ Pas de graphiques ☐ Les tableaux ou les graphiques ne sont pas nécessaires ou convenables pour le rapport	☐ tableaux efficaces ☐ graphiques efficaces	Médiocre 1 2 Adéquate 3 4 Excellente 5 S.O. 9		
	9.2.6 Bien organiser le rapport pour qu'il soit facile à consulter				Médiocre 1 2 Adéquate 3 4 Excellente 5		
	9.2.7 L'examen est entravé par le degré d'interdiction en vertu de la Loi sur l'accès à l'information		☐ Oui – grandement ☐ Oui – légèrement ☐ Non				
10. Évaluation globale							
	10.1 Évaluation globale				Médiocre 1 2 Adéquate 3 4 Excellente 5		

Annexe B
Répartition des rapports examinés par ministère et organisme

Répartition des rapports examinés par ministère et organisme

Ministère et organisme	Nombre de rapports
Affaires étrangères et Commerce international	3
Affaires indiennes et du Nord Canada	10
Agence canadienne de développement international	4
Agence de développement économique du Canada pour les régions du Québec	3
Agence de promotion économique du Canada atlantique	1
Agence des douanes et du revenu du Canada	2
Agence spatiale canadienne	1
Agriculture et Agroalimentaire Canada	3
Anciens Combattants Canada	2
Bureau de la protection des infrastructures essentielles et de la protection civile	1
Centre canadien d'hygiène et de sécurité au travail	1
Centre canadien de gestion	1
Citoyenneté et Immigration Canada	1
Commission de la fonction publique	1
Commission nationale des libérations conditionnelles	1
Communications Canada/Bureau d'information du Canada	3
Condition féminine Canada	1
Conseil de recherches en sciences naturelles et en génie	1
Conseil national de recherches du Canada	5
Défense nationale	2
Défense nationale/Anciens Combattants Canada	1
Développement des ressources humaines Canada	5
Finances Canada	1
Gendarmerie royale du Canada	1
Industrie Canada	12
Instituts de recherche en santé du Canada	2
Justice Canada	4
Ministère de la Diversification de l'économie de l'Ouest	5
Patrimoine canadien	11
Pêches et Océans Canada	2
Ressources naturelles Canada	10
Santé Canada	6
Secrétariat du Conseil du Trésor	2
Service correctionnel du Canada	3
Transports Canada	3
Total	115